DOCUMENT RESUME

ED 480 275 EA 032 726

AUTHOR	Kellaghan, Thomas; Greaney, Vincent
TITLE	Using Assessment To Improve the Quality of Education. Fundamentals of Educational Planning.
INSTITUTION	United Nations Educational, Scientific, and Cultural Organization, Paris (France). International Inst. for Educational Planning.
SPONS AGENCY	Swedish International Development Authority (SIDA).
REPORT NO	No-71
ISBN	ISBN-92-803-1215-4
PUB DATE	2001-00-00
NOTE	99p.
AVAILABLE FROM	United Nations Educational, Scientific, and Cultural Organization, International Inst. for Educational Planning, 7, place de Fontenoy, 75352 Paris 07 SP, France. Web site: http://www.unesco.org/ iiep.
PUB TYPE	Books (010) -- Guides - Non-Classroom (055)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS
DESCRIPTORS	Academic Achievement; Academic Standards; *Accountability; Decision Making; *Educational Assessment; Educational Improvement; Educational Objectives; *Educational Planning; *Educational Quality; Elementary Secondary Education; *Evaluation Utilization; Foreign Countries; International Education; National Standards; School Effectiveness; State Standards

ABSTRACT

The purpose of this booklet, which draws on experience from all over the world, is to provide educational planners, administrators, and managers with analysis and information on using assessment to improve the quality of education. Assessment options and what research and other kinds of evidence reveal about their advantages and disadvantages are described. Consideration is also given to the effects of assessment on curricula, teaching, and learning. The focus of the booklet is on national and international assessments of student achievement, since these are regarded as important sources of information for education policymakers. Following the introduction, definitions of key concepts in current reform proposals are presented in chapter 2. Chapter 3 outlines characteristics of the kind of assessment that is proposed, with a view toward improving quality. In chapter 4, national and state assessments of student achievement are described. Chapter 5 focuses on various ways of variety of ways of describing student achievement, as well as ways in which information derived from national assessments. International assessments of student achievement are the subject of chapter 6. (Author)

Reproductions supplied by EDRS are the best that can be made
from the original document.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

J. Hall

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Fundamentals of Educational Planning

ED 480 275

Using assessment to improve the quality of education

Thomas Kellaghan
Vincent Greaney

UNESCO:
International
Institute
for Educational
Planning

EA 032 736

BEST COPY AVAILABLE

2

Included in the series:*

2. The relation of educational plans to economic and social planning, *R. Poignant*
4. Planning and the educational administrator, *C.E. Beeby*
5. The social context of educational planning, *C.A. Anderson*
6. The costing of educational plans, *J. Vaizey, J.D. Chesswas*
7. The problems of rural education, *V.L. Griffiths*
8. Educational planning: the adviser's role, *A. Curle*
9. Demographic aspects of educational planning, *Ta Ngoc C.*
10. The analysis of educational costs and expenditure, *J. Hallak*
11. The professional identity of the educational planner, *A. Curle*
12. The conditions for success in educational planning, *G.C. Ruscoe*
13. Cost-benefit analysis in educational planning, *M. Woodhall*
18. Planning educational assistance for the second development decade, *H.M. Philips*
20. Realistic educational planning, *K.R. McKinnon*
21. Planning education in relation to rural development, *G.M. Coverdale*
22. Alternatives and decisions in educational planning, *J.D. Montgomery*
23. Planning the school curriculum, *A. Lewy*
24. Cost factors in planning educational technological systems, *D.T. Jamison*
25. The planner and lifelong education, *P. Furter*
26. Education and employment: a critical appraisal, *M. Carnoy*
27. Planning teacher demand and supply, *P. Williams*
28. Planning early childhood care and education in developing countries, *A. Heron*
29. Communication media in education for low-income countries, *E.G. McAnany, J.K. Mayo*
30. The planning of nonformal education, *D.R. Evans*
31. Education, training and the traditional sector, *J. Hallak, F. Caillods*
32. Higher education and employment: the IIEP experience in five less-developed countries, *G. Psacharopoulos, B.C. Sanyal*
33. Educational planning as a social process, *T. Malan*
34. Higher education and social stratification: an international comparative study, *T. Husén*
35. A conceptual framework for the development of lifelong education in the USSR, *A. Vladislavlev*
36. Education in austerity: options for planners, *K. Lewin*
37. Educational planning in Asia, *R. Roy-Singh*
38. Education projects: elaboration, financing and management, *A. Magnen*
39. Increasing teacher effectiveness, *L.W. Anderson*
40. National and school-based curriculum development, *A. Lewy*
41. Planning human resources: methods, experiences and practices, *O. Bertrand*
42. Redefining basic education for Latin America: lessons to be learned from the Colombian Escuela Nueva, *E. Schiefelbein*
43. The management of distance learning systems, *G. Rumble*
44. Educational strategies for small island states, *D. Atchoarena*
45. Judging educational research based on experiments and surveys, *R.M. Wolf*
46. Law and educational planning, *I. Birch*
47. Utilizing education and human resource sector analyses, *F. Kemmerer*
48. Cost analysis of educational inclusion of marginalized populations, *M.C. Tsang*
49. An efficiency-based management information system, *W.W. McMahon*
50. National examinations: design, procedures and reporting, *J.P. Keeves*
51. Education policy-planning process: an applied framework, *W.D. Haddad, with the assistance of T. Demsky*
52. Searching for relevance: the development of work orientation in basic education, *W. Hoppers*
53. Planning for innovation in education, *D.E. Inbar*
54. Functional analysis (management audits) of the organization of ministries of education, *R. Sack, M. Saïdi*
55. Reducing repetition: issues and strategies, *T.O. Eisemon*
56. Increasing girls and women's participation in basic education, *N.P. Stromquist*
57. Physical facilities for education: what planners need to know, *J. Beynon*
58. Planning learner-centred adult literacy programmes, *S.E. Malone, R.F. Arnove*
59. Training teachers to work in schools considered difficult, *J.-L. Auduc*
60. Evaluating higher education, *J.L. Rontopoulou*
61. The shadow education system: private tutoring and its implication for planners, *M. Bray*
62. School-based management, *I. Abu-Duhou*
63. Globalization and educational reform: what planners need to know, *M. Carnoy*
64. Decentralization of education: why, when, what and how?, *N. McGinn, T. Welsh*
65. Early childhood education: need and opportunity, *D. Weikart*
66. Planning for education in the context of HIV/AIDS, *M.J. Kelly*
67. Legal aspects of educational planning and administration, *C. Durand-Prinborgne*
68. Improving school effectiveness, *J. Scheerens*
69. Reviewing quantitative research to inform policy processes, *S.J. Hite*
70. National strategies for e-learning in post-secondary education and training, *T. Bates*

* Also published in French. Other titles to appear.

Using assessment to improve the quality of education

Thomas Kellaghan and Vincent Greaney

Paris 2001
UNESCO: International Institute for Educational Planning

The Swedish International Development Co-operation Agency (Sida) has provided financial assistance for the publication of this booklet.

Published in 2001 by the United Nations
Educational, Scientific and Cultural Organization
7 place de Fontenoy, F 75352 Paris 07 SP
Printed in France by Imprimerie Alençonnaise
Cover design by Pierre Finot

ISBN 92-803-1215-4
© UNESCO 2001

Fundamentals of educational planning

The booklets in this series are written primarily for two types of clientele: those engaged in educational planning and administration, in developing as well as developed countries; and others, less specialized, such as senior government officials and policy-makers who seek a more general understanding of educational planning and of how it is related to overall national development. They are intended to be of use either for private study or in formal training programmes.

Since this series was launched in 1967 practices and concepts of educational planning have undergone substantial change. Many of the assumptions which underlay earlier attempts to rationalize the process of educational development have been criticized or abandoned. Even if rigid mandatory centralized planning has now clearly proven to be inappropriate, this does not mean that all forms of planning have been dispensed with. On the contrary, the need for collecting data, evaluating the efficiency of existing programmes, undertaking a wide range of studies, exploring the future and fostering broad debate on these bases to guide educational policy and decision-making has become even more acute than before. One cannot make sensible policy choices without assessing the present situation, specifying the goals to be reached, marshalling the means to attain them and monitoring what has been accomplished. Hence planning is also a way to organize learning: by mapping, targeting, acting and correcting.

The scope of educational planning has been broadened. In addition to the formal system of education, it is now applied to all other important educational efforts in non-formal settings. Attention to the growth and expansion of education systems is being complemented and sometimes even replaced by a growing concern for the quality of the entire educational process and for the control of its results. Finally, planners and administrators have become more and more aware of the importance of implementation strategies and of the role of different regulatory mechanisms in this respect: the choice of financing methods, the examination and certification procedures or various other regulation and incentive structures. The concern of planners is twofold: to reach

a better understanding of the validity of education in its own empirically observed specific dimensions and to help in defining appropriate strategies for change.

The purpose of these booklets includes monitoring the evolution and change in educational policies and their effect upon educational planning requirements; highlighting current issues of educational planning and analyzing them in the context of their historical and societal setting; and disseminating methodologies of planning which can be applied in the context of both the developed and the developing countries.

For policy-making and planning, vicarious experience is a potent source of learning: the problems others face, the objectives they seek, the routes they try, the results they arrive at and the unintended results they produce are worth analysis.

In order to help the Institute identify the real up-to-date issues in educational planning and policy-making in different parts of the world, an Editorial Board has been appointed, composed of two general editors and associate editors from different regions, all professionals of high repute in their own field. At the first meeting of this new Editorial Board in January 1990, its members identified key topics to be covered in the coming issues under the following headings:

1. Education and development.
2. Equity considerations.
3. Quality of education.
4. Structure, administration and management of education.
5. Curriculum.
6. Cost and financing of education.
7. Planning techniques and approaches.
8. Information systems, monitoring and evaluation.

Each heading is covered by one or two associate editors.

The series has been carefully planned but no attempt has been made to avoid differences or even contradictions in the views expressed by the authors. The Institute itself does not wish to impose any official doctrine. Thus, while the views are the responsibility of the authors and may not always be shared by UNESCO or the IIEP, they warrant attention in the international forum of ideas. Indeed, one of the purposes

of this series is to reflect a diversity of experience and opinions by giving different authors from a wide range of backgrounds and disciplines the opportunity of expressing their views on changing theories and practices in educational planning.

Assessment has become over the years an important key to the improvement of the quality of education. It is one of the most reliable ways of identifying problems, whether these are at the system level, the school level, or concern the individual student.

Nevertheless, assessment raises a certain number of questions. Opinions do not always converge on what exactly should be measured, or in what way it should be measured. And, for a variety of reasons, it is often difficult to measure the evolution of quality over time.

Another critical issue is what should be done with the results of an assessment. There is much debate around whether and how these should be used for accountability purposes.

Thomas Kellaghan and Vincent Greaney tackle these and many other questions, and show that there are no easy answers. Policy-makers will find this thought-provoking booklet a useful guide in defining and implementing assessment strategies.

Gudmund Hernes
Director, IIEP

Composition of the Editorial Board

Chairman: Gudmund Hernes
Director, IIEP

General Editors: Françoise Caillods
IIEP

T. Neville Postlethwaite
(Professor Emeritus)
University of Hamburg
Germany

Associate Editors: Jean-Claude Eicher
University of Bourgogne
France

Claudio de Moura Castro
International Consultant
Brazil

Kenneth N. Ross
IIEP

Richard Sack
International Consultant
France

Rosa Maria Torres
International Consultant
Argentina

Preface

How to introduce substantial educational change and improve quality has been the concern of educational planners for many years now. This concern with quality has progressively shifted its focus from input to outcomes in terms of learning achievement.

Past educational reforms encompassing structure, curriculum and teacher training have to a large extent given way to more 'modest' reforms, oriented towards the improvement of learning achievement, school effectiveness, management and accountability. Decentralization, school-based management and learning assessment became key concepts in the educational reforms of the 1990s. Global economic competition has brought to the fore the critical importance of quality of human resources, and the demand for new competencies in today's information society. The educational system, schools and individual students are all under increasing pressure to perform.

Assessment and particularly the assessment of students' learning achievements have become the object of a good deal of attention and activities all over the world, in industrialized countries and developing countries alike.

International surveys comparing the achievement of students in literacy, mathematics and science in different countries are flourishing. The publication of the results of such surveys is a major event in most participating countries, often to the detriment of a more thorough analysis of the differences in achievement between schools and between groups, and the factors explaining such differences. Most countries are also embarking upon regular assessments of their students. Such studies are not only intended to help national authorities to monitor the system, they are considered as a lever of change in itself, likely to bring about changes in schools and classrooms.

In view of the present emphasis on assessment, educational policy-makers are often faced with the question of what sort of assessment to introduce and for what purpose. The present booklet, written by

Thomas Kellaghan, Director of the Educational Research Centre at St Patrick's College in Dublin, and Vincent Greaney, Lead Education Specialist at the World Bank, will answer many of their questions. They clarify basic concepts such as assessment, quality, standards and accountability, but also present the characteristics of assessment in several reform proposals, discuss various models of national assessment, and raise critical issues in the design of an assessment.

While at the same time recognizing that national assessments are not the only way of monitoring performance and enhancing quality, the authors present and discuss in an extremely clear and concise way the most important factors to be taken into account when reflecting upon the use of assessment within a specific national context. The Editorial Board is very grateful to Thomas Kellaghan and Vincent Greaney for this stimulating and topical piece.

Françoise Caillods
Co-General Editor

Acknowledgements

This booklet grew out of a paper that was delivered by the first-named author at a meeting of the International Working Group on Education in Florence, 14-16 June 2000. The paper was prepared at the suggestion of Liz Higgins of Ireland Aid; preparation of the paper and the author's attendance at the meeting were funded by Ireland Aid.

The authors wish to express their appreciation to Dr Françoise Caillods of the International Institute for Educational Planning for her suggestion that the booklet be prepared, and for her support and patience during its longer-than-expected gestation; to Professor T. Neville Postlethwaite for his input and direction at an early stage of the work; to colleagues at the World Bank, in particular Luis Benveniste, for their assistance; and to Hilary Walshe and Regina Klöditz who typed the manuscript.

Contents

Preface		9
Acknowledgements		11
I.	Introduction	15
II.	Assessment, quality, standards, and accountability	19
	Assessment in education	19
	Quality in education	22
	Standards	23
	Accountability	26
	Conclusion	27
III.	Characteristics of assessment in current reform proposals	29
	Focus on outcomes	29
	Focus on external assessment	31
	Assessment as a lever of reform	31
	Conclusion	32
IV.	National assessments of student achievement	33
	What is a national assessment?	33
	Models of national assessment	36
	Why carry out a national assessment?	39
	Information needs addressed in a national assessment	41
	Growth in national assessment activity	42
	Conclusion	44
V.	Describing and using information derived from a national assessment	47
	Describing student achievements	47
	The communication of information derived from a national assessment	55
	The use of national assessment data	58
	Conclusion	60
VI.	International assessments of achievement	63
	What is an international assessment?	63

	Why participate in an international assessment?	64
	Issues in international assessments	65
	The uses of international assessment data	68
	Conclusion	69
VII.	Using student achievement data to evaluate schools	71
	Evaluating the quality of schools	71
	Action following the evaluation of schools	72
	League tables	75
	The impact of national assessment data on schools	79
	Conclusion	81
VIII.	Issues in the design of an assessment	83
	Considerations in designing an assessment	85
	Participation in an international assessment	88
	Cost	88
	The use of information derived from an assessment	89
	Conclusion	91
References		93

I. Introduction

The need to improve the quality of education is voiced frequently in contemporary writings, in political debate, and in the media. It is a need that is just as likely to be articulated in a country with a well-developed education system in the industrialized world, as in a developing country that has not yet reached the goal of universal primary education. Concern about educational quality is not new, of course, and over the years, various reforms have been proposed in response to it, ranging from changes in the organization of schools and curricula, to improved management structures in schools, teacher education, peer tutoring, and greater parental involvement. Reforms continue in these areas today; however, the reform that has dominated the education scene in the 1980s, 1990s, and into the 2000s, proposes the use of assessment to improve educational quality.

In considering this reform, and proposals to implement it, education policy-makers, planners, and managers are faced with two major decisions. The first is: should resources be invested in assessment activities (either to improve existing ones or to introduce new ones), or would better results be achieved if resources were invested in some other aspect of the education system, such as school buildings, textbooks, teacher training, or the number of teachers? Given the contemporary worldwide emphasis on assessment, it is unlikely that it will be dismissed as an area that does not merit further consideration and investment, in which case a further question arises. What kind of assessment, and what uses of assessment information, are likely to impact on the quality of education in schools? There are, of course, no simple answers to these questions. Answers will be arrived at only when all relevant factors (including needs for information, value positions, financial, political and logistic considerations, and traditions of assessment) have been taken into account. Conceptual analysis of relevant issues and the availability of information on practice in other education systems can also be of assistance in evaluating options and in anticipating the effects of proposed actions.

It is the purpose of this booklet, which draws on experience from all over the world, to provide educational planners, administrators, and managers with such analysis and information. Assessment options, and what research and other kinds of evidence reveal about their advantages and disadvantages, are described. Further, since assessment is being proposed not just to describe achievement, but as a lever of change in education systems leading to improved quality and higher standards of learning, consideration will be given to the effects of assessment on curricula, teaching, and learning.

The focus of the booklet is on national and international assessments of student achievement since these (in particular, the former) are regarded as important sources of information for education policy-makers and managers. A focus on these areas is not to devalue other areas of assessment. For example, the assessment of students' learning in the classroom, by their own teachers, merits serious consideration in the context of improving the quality of education. However, classroom assessment has not figured largely in reform proposals, and will not be considered as a separate topic in this booklet. A further type of assessment that will not be considered is the external or public examination, even though its effects on classroom practice have been well documented, and, indeed, some commentators have argued that well-designed examinations also have an important role to play in improving student learning.

Following the introduction, definitions of key concepts in current reform proposals (assessment, quality, standards, and accountability) are presented in *Chapter II*. In *Chapter III*, characteristics of the kind of assessment that is being proposed with a view to improving quality are outlined. In *Chapter IV*, national (and state) assessments of student achievement are described. While relatively new on the education scene, these assessments have shown remarkable growth in the 1990s. They are the focus of much discussion and activity, and are the main types of assessment associated with quality and standards in current debate. A variety of ways that have been used to describe student achievement, as well as ways in which information derived from national assessments has been used are described in *Chapter V*. International assessments of student achievement, which are the

subject of *Chapter VI*, also grew in scale in the 1990s. They use procedures that are very similar to those used in national assessments. However, since they have to meet agreed common needs of participating countries, compromise will be required in their execution (relating, for example, to the design of assessment instruments), and the assessments will be less attuned than national assessments to the specific needs of individual countries. They do, however, allow for comparisons between the performances of students in different education systems, and so have extended interest in standards to an international context, generating much media and political interest.

Although the evaluation of schools has traditionally been carried out by school inspectors or supervisors in many countries, a number of countries in the 1990s began to use the performance of students on tests of achievement in national and state-wide assessments for this purpose. This use, and problems associated with it, are described in *Chapter VII*. In the final chapter, decisions involved in the design of an assessment, and the way information derived from an assessment might be used, are described.

II. Assessment, quality, standards, and accountability

To set proposals to use assessment to improve the quality of education in context, a number of key concepts are described in this chapter. The meanings, forms, and purposes of assessment are outlined, following which three associated concepts are considered: quality, standards, and accountability.

Assessment in education

Definition of assessment

The term 'assessment' may be used in education to refer to any procedure or activity that is designed to collect information about the knowledge, attitudes, or skills of a learner or group of learners. A more detailed definition is that assessment is:

> the process of obtaining information that is used to make educational decisions about students, to give feedback to the student about his or her progress, strengths, and weaknesses, to judge instructional effectiveness and curricular adequacy, and to inform policy (AFT, NCME, NEA, 1990, p.1).

In the type of assessment that everyone who has been to school will recognize, a teacher or examiner is involved. An assessment could also be carried out by the learner himself or herself, or by another party. Whatever the circumstances, assessment, as reflected in the above definition, has traditionally been associated with the appraisal of individuals. The term, however, has been used with increased frequency since the 1980s to describe the performance of schools or of school systems.

Forms of assessment

There are many ways in which an assessment can be carried out. It may be oral, aural, written, practical, or observational. It can be

carried out in one-to-one, small group, or large group settings, or on a computer. If a holistic view of achievement is adopted, assessment will be based on students' complete performance on a complex task (e.g. constructing a wooden object, arguing a position in an extended essay, cooking a meal). If, on the other hand, an analytic view, in which achievement is broken down into a series of small chunks, is adopted, assessment will comprise a variety of short-answer responses to questions. Degree of formality may also vary, from one in which an individual responds to a series of questions or prompts under controlled conditions within a limited time-frame, to a situation in which an individual's behaviour is observed in a non-intrusive way over an extended period of time.

Purposes of assessment

Assessment information is used for a variety of purposes. At the individual student level, it is used (a) to describe students' learning, to identify and diagnose learning problems, and to plan further teaching/learning; (b) to provide guidance for students in selecting further courses of study or in deciding on vocational options; (c) to motivate students by providing goals or targets, by clarifying the nature of learning tasks, and by letting students, and their teachers, know how they are progressing; (d) to certify that individuals have reached a certain level of competence; and (e) to select individuals for the next level of the education system or for a job.

The main purposes of an assessment of an institution or system are to reach a judgement about the effectiveness of a school, and to reach a judgement about the adequacy of the performance of an education system or of a part of it (see *Box 2.1*). Even in these cases, assessment is based on the performances of individual students, which are aggregated to the level of the institution or system, usually to provide information for policy-makers and others.

Box 2.1 The use of assessment information to evaluate schools and systems

School evaluation. The results of an assessment are sometimes used as a measure of the effectiveness of a school. 'Successful' performance may be equated with the percentage of students who reach a given level of 'proficiency' or the percentage that advance to the next highest level in the education system. In some countries, the process is made public by the publication of results in league tables, or by individual schools publicizing academic results in newspaper advertisements.

System evaluation. Assessment information may be used to reach a judgement about the adequacy of the performance of an education system or of a part of it. The standard implied in the judgement might be predetermined (e.g. 80 per cent of students achieve a specified level of proficiency) or it might be comparative (e.g. performance at a previous point in time or performance of students in a similar system).

While a distinction can be made between the purposes of individual, institutional, and system assessments, in practice it may be blurred. Some assessments are designed with several purposes, and are being used to provide information, not only on education systems, but also on schools, and even to perform functions that were traditionally associated with assessment at the individual student level (e.g. relating to grade promotion, motivation, and certification).

Considerations when using assessment information

It is easy to give the impression in describing the results of an assessment that the definition and measurement of human achievements are unproblematic. Reports in the media, in particular, seem to assume that assessments provide information that is as unambiguous, objective, and reliable as information about physical characteristics such as height or weight.

It is important that those involved in making decisions based on assessment information do not adopt this view, and are sensitive to difficulties in defining and measuring achievement. One difficulty is that an assessment task can include only a sample of relevant

knowledge and skills, and a sample of student responses. In this situation, it is reasonable to ask to what extent performance on a task, especially if it is limited to written responses under controlled conditions, can reasonably be generalized to other tasks and settings. Furthermore, if one considers the kinds of achievement (knowledge and skills) that reformers say are important in contemporary society for ensuring economic competitiveness, one may ask, for example, how investigative skills or a commitment to learning are to be assessed? To take a concrete example from mathematics, how is the ability to construct and apply principles (rather than just demonstrate computational procedures) or the ability to use a variety of methods to solve non-routine problems assessed?

Apart from technical problems in measuring achievement, it should also be recognized that assessment is a social-political activity. This will be the case when assessment procedures are defined and mandated in policy decisions. Furthermore, the information derived from an assessment can have far-reaching personal and social consequences, intended and unintended, positive and negative, especially when used to make decisions about students' educational careers, or to hold individuals or institutions accountable.

Quality in education

In much contemporary discourse on education, the word *quality* is frequently mentioned, although it is rarely defined. In the absence of an agreed formal definition, it may be inferred from normal usage that the term relates to the adequacy or appropriateness of objects or processes for the purposes for which they were intended. Furthermore, quality implies a scale, and often implies standards. An object can be of 'good' or 'poor' quality, or it can meet, or fail to meet, a standard.

A problem in defining the quality of education arises when one chooses the aspect of education that will be the focus of attention. Since education has many purposes and components, questions regarding quality may reasonably be posed about any important aspect of a system: infrastructure, school buildings, administration, teacher training, educational materials, teaching, or student achievements. All

these elements, it will be noted, are interrelated, and a serious deficit in one is likely to have implications for quality in others.

If the focus is on the outcomes of education (as it is in current reform proposals), a further problem arises since there is not general agreement on what the purposes of schooling should be. Purposes or objectives are matters of choice, and so must be based on value judgments. For some, the role of education is fostering students' cognitive, moral, and social development; for others, education is a means of promoting social cohesion and nation building; for others, it is a preparation for the world of work. It may be because of this divergence in views that many, though not all, system assessments focus on knowledge and skills that are universally accepted as important. Can students read? Can they compute? Have they an understanding of basic scientific concepts? This focus, however, should not lead to a neglect of the higher order cognitive skills that many reforms aspire to promote, or of many other valued outcomes of schooling relating to attitudes, values, aspirations, motivation, self-concept, and socialization.

Standards

The idea of standards is closely related to the idea of quality, and has had a part in much debate about education since the 1980s. Indeed, concern for standards was at the heart of the political debate in the USA that resulted in proposals for reform involving assessment. Critics of the education system concluded, often relying on evidence from national and international assessments that was less than adequate, that students were not being taught, and thus were not acquiring, the knowledge and skills that they would need to compete economically and to sustain a vibrant democracy.

In considering what is meant by a standard, a number of dictionary definitions seem relevant to education: degree of excellence required for a particular purpose; a measure of what is adequate; and a socially or practically described level of performance. Of these three, little attention has been given to the first in defining educational standards. Most efforts have been more limited in scope, and have focused on the knowledge and skills that are considered appropriate and adequate

for students to have acquired at particular levels in the education system. Following the introduction of the Revised Code into elementary schools in Great Britain in 1862, a distinctly educational meaning associated with assessment was assigned to the word 'standard', when it was defined in the *Oxford English Dictionary* as "each of the recognized degrees of proficiency, as tested by examination, according to which schoolchildren are classified". This meaning fell into disuse when central examinations were abolished, although the term continued in use for some time as a grade of classification in primary schools.

Today, the issue of standards in education is as prominent as it was in the nineteenth century. The United Kingdom has a School Standards and Framework Act (1998) and a Standards and Effectiveness Unit in its Department for Education and Employment. In the United States, criticisms of the quality of education led to proposals to adopt standards relating to the curriculum content or framework appropriate for grade levels (*content standards*) and to the achievements expected of students at the grade levels (*performance standards*). Underlying the proposals was a perceived need for greater uniformity in what was taught in schools throughout a country that did not have a national curriculum and in which education was controlled by state, not federal, government. Many countries will not have to address this issue, since they will have long-standing national curricula, in many cases reinforced by an external examination system.

Other aspects of American proposals, however, merit attention, even though they often seem to be disregarded when reforms based on the use of assessment are introduced in other countries. First, the standards specify challenging definitions of proficiencies, which have been variously described as comprising higher-order thinking skills, problem-solving abilities, investigative and reasoning skills, improved means of communication, and a commitment to lifelong learning. Secondly, the standards are intended for *all* students. Groups of students will not be written off, and offered less challenging curricula. Thirdly, standards constitute only one element of reform, and their adoption is unlikely to have much impact unless all major instructional guidance mechanisms in the education system are aligned with them: student assessments, instructional materials, teacher preparation and

later professional development, textbooks, school capacity building and improvement, supportive services to schools, and equity targeting devices. In particular, it is recognized that students will not develop the required proficiencies unless there are changes in curricula and classrooms. New approaches will be required, involving greater student engagement; learning that is heuristic and based on experience in concrete situations, rather than axiomatic and abstract; and the provision of opportunities for students to develop strategies for problem-solving, thinking effectively, and learning for themselves. The particular role of standards will be to help make concrete for teachers the goals of a system, and to assist them and others in focusing on what students at varying levels of achievement should know and be able to do.

The setting of standards for student performance involves identifying one or more threshold scores on a total mark scale, which define levels or grades corresponding to the intervening range of scores. The process, however, is not without its problems. First, the achievements of a group of students do not fall into neat discrete categories, but rather form a continuum in which there are very small differences between the scores of individual students. There is no obvious basis for deciding on the point at which a standard can be said to be satisfied, or for saying that students who score just above a dividing point differ substantially from students who score just below it in their achievements. Second, a summary score or grade can be achieved through different patterns of responses. Thus, the knowledge and skills that merit an identical grade can differ from student to student. Third, there are no absolute criteria against which the results of a standard-setting procedure can be judged. Although standards may be proposed on the basis that it is important to provide students with the knowledge and skills that they will need for later life, it is not possible to specify these precisely and, anyhow, they will differ from student to student depending on the kind of work a student will eventually do.

Fourth, the actual setting of standards is usually based on the judgements of individuals. For this reason they are regarded as arbitrary, not however in the sense that they are capricious or that they are not based on reason, but because they involve the exercise of choice and

the application of values. While the individuals who make the judgements will be chosen because of their knowledge of the content of curricula and/or about the students for whom an assessment is designed, and while various procedures are available to help achieve consensus between individuals, the judgements will always be subject to human error. Fifth, standard setting will be based, explicitly or implicitly, on policy decisions, when, for example, the emphasis is placed on 'minimum competency' or on 'excellence'. As with all social-political decisions, these decisions will be matters of judgement, will be subject to change, and will involve trade-offs between competing values and goals (Kane, 2001).

Even if standards are highly prescribed and detailed, and are accompanied by centralized assessment, proposals to use them to specify teacher behaviour have to confront the realities of the classroom. It cannot be assumed that all teachers in all schools will teach the same content and skills. The work of individual teachers will always be influenced by a variety of other factors: their understanding of reform proposals and, in particular, of what the standards represent; their prior knowledge and experience; the conditions in which they teach; and the characteristics of their students.

Accountability

Proposals to use assessment to improve standards in schools owe much to accountability movements which, in response to political, social, and economic pressures, are now a feature of government administrations in many countries (Kirst, 2000). If quality does not reach an acceptable standard, it seems reasonable to ask: who is accountable?

Accountability may be defined as the state of being responsible and answerable for the discharge of duties. When one examines how the concept has been applied in the field of education, and is used in contemporary debate, two issues are clear. First, accountability is focused for the most part on specified results or outcomes. This is so, despite the fact that it might be expected that account would be taken of all that goes on in schools, not just student achievements, requiring,

for example, schools to explain themselves in terms of their mission, organization, activities, and priorities for students (Haney and Raczek, 1994).

Secondly, accountability systems are often silent about the responsibilities of individuals or agencies other than teachers and schools. While it is obvious that teachers and students should bear some responsibility for learning, it would seem reasonable that they should be held accountable only for matters over which they have control, and that the responsibility of other institutions, agencies, and individuals that exercise control over the resources and activities of schools should also be reflected in an accountability system. Thus, government (central and local, depending on how the education system is organized), educational planners, and managers, would seem accountable for some aspects of education. Others who are not directly involved in the administration of the system or in the operation of schools might also be regarded as accountable to some degree (e.g. teacher-training institutions, parents, and even taxpayers). If, for example, an evaluation of the adequacy of existing resources pointed to deficiencies in the curriculum or in teacher preparation, the responsibility for dealing with these could hardly be laid at the door of individual teachers. Similarly, if inequities in the distribution of resources (e.g. in the availability of science courses for female students), or the need to target some categories of student or certain schools for special treatment, are identified, the accountability of those responsible for the provision of resources would arise. Assessment systems differ in the extent to which teachers and others involved in the educational process are held accountable for student achievements.

Conclusion

In this chapter, key concepts in proposals to use assessment to improve educational quality – assessment, quality, standards, and accountability – were defined. When these terms are used in reform rhetoric, they are often presented in a fairly simplistic way, and as relatively unproblematic. The brief review presented in this chapter raises questions about this use, and should serve to underline the complexity of the issues involved. It should also point to the need to

consider assessment in the context of other elements of the education system when its potential to influence the quality of student learning is being evaluated.

III. Characteristics of assessment in current reform proposals

A concern with quality immediately gives rise to the issue of assessment, since some form of appraisal of the elements of an education system is necessary before a judgement can be made about its quality. Current reform movements go further, and specify three characteristics of the kind of assessment that is likely to impact on quality. First, the focus is on outcomes. Secondly, the assessment that is conducted (and perhaps mandated) involves an agency outside the school. And thirdly, the expectation is that assessment will act as a lever of reform.

Focus on outcomes

In the past, though there have been exceptions, the main focus in assessing the quality of education was on inputs (e.g. student participation rates, physical facilities, curriculum materials, books, and teacher training). This is no longer the case, deficient though any of these might be in many countries. Today, the question posed by many commentators and policy-makers focuses on the *outcomes* of education: are students, as a result of their exposure to schooling, acquiring appropriate knowledge, skills, behaviour, and attitudes? This concern is understandable, given that systematic objective information about student learning in education systems is fairly rare, although the need for it for a variety of purposes is clear. Policy-makers or managers, for example, may need information that would help them reach a judgement about the adequacy of student achievements generated in the system; or they may need baseline data on achievements against which to measure progress in an educational reform programme, or data that would allow them to document demographic differences. Teachers may feel the need for information about achievements in the system with which to compare the achievements of their own students or to assess their effectiveness in teaching.

One of the most influential statements of concern for outcomes is to be found in the document adopted by the World Conference on Education for All in Jomtien, Thailand in March, 1990, which stated that the provision of basic education for all was meaningful only if children actually acquired useful knowledge, reasoning ability, skills, and values. As a consequence, Article 4 of the *World Declaration on Education for All* (1990) stated that the focus of basic education should be "on actual learning acquisition and outcome, rather than exclusively upon enrolment, continued participation in organized programmes and completion of certification requirements" (p. 5). The Dakar Framework for Action (2000), produced at the end of the 10-year follow-up to Jomtien, again stressed the importance of having "a clear definition and accurate assessment of learning outcomes (including knowledge, skills, attitudes, and values)" as governments work to ensure basic education of quality for all (UNESCO, 2000).

Four issues can be identified as giving rise to concern about outcomes. First, many students (particularly ones with short education careers) seem to benefit little from their educational experiences. In a study in Bangladesh, for example, only slightly over one-third of those who had just completed primary school were found to have achieved a minimum competency level in basic learning skills (reading, writing, and mathematics) (Greaney, Khandker and Alam, 1998). Second, quality is perceived to be deteriorating in many countries, although evidence to support this perception is usually not available. However, it is argued that deterioration is inevitable in a situation in which fewer resources per student are available, as financial inputs decrease and student numbers increase. Third, given the importance of schooling in economic reform and the need to prepare students to meet the demands of the workplace, there is concern that the competences acquired even by students who stay in the education system for a long time may not be adequate to meet the needs of the information-based global competitive economy of the twenty-first century. Finally, education ministries, like other government ministries, are increasingly seeking evidence relating to the outcomes of their substantial investments.

Focus on external assessment

A second feature of current reforms involving assessment is that most resources have been invested in assessment that is carried out by an agent outside the school. There are a number of reasons for this. First, such assessment can be externally mandated; one does not have to rely totally on schools to implement it. Second, a programme of assessment can be established and implemented in a relatively short period of time. Third, such a programme is inexpensive compared, for example, to reducing class size or implementing curriculum changes that might involve substantial professional development for teachers. Fourth, pressure can be placed on schools to improve by making the results of an assessment visible, even reporting them in the press. Fifth, policy-makers can reasonably expect some improvement in test scores over time as a result of increased attention in schools to the areas covered in the assessment. However, this improvement will not necessarily be matched by an improvement in the broader achievement constructs (e.g. problem-solving ability, investigative skills) that the assessments were intended to measure (Linn, 1983, 2000; Torrance, 1997).

Finally, in many countries, education authorities have been granting greater freedom to individual schools to manage their own affairs, or, as a result of being constantly required to negotiate with unions even small changes in practice, have been losing control over what goes on in schools. The erosion of the role of school inspectorates illustrates this trend. In these situations, technological advances (particularly ones associated with assessment and the processing of large amounts of data expeditiously and inexpensively) have provided the means of establishing an alternative system of accountability based on students' achievements.

Assessment as a lever of reform

A third feature of current reforms is that assessment is considered to be not just a means of obtaining information about education systems, but also a lever of reform (Madaus and Kellaghan, 1992; Popham, 1987). Thus, assessment is ascribed an active role, one in

which it is consciously used in the educational process to change teacher behaviour and classroom instruction which, in turn, is expected to raise the standard of students' learning.

The use of assessment as a lever of change gives rise to a number of issues. First, assessment policy, if it is to be effective, should be linked to, and consistent with, policy relating to other instructional guidance mechanisms in the education system (e.g. curricula, other assessment procedures, teacher education, textbooks). Second, information based on an assessment should be made available without undue delay in a readily understandable form, certainly to educational planners and administrators, probably to schools, and possibly to the general public. Third, the issue of accountability (as noted in *Chapter II*) will arise when action on the basis of information is considered. Two options present themselves. A ministry of education might accept primary responsibility for standards and use assessment information to make decisions about the allocation of resources. On the other hand, schools might be held primarily responsible, with the expectation that they would take whatever action was required.

Conclusion

Assessment has many meanings and uses. Current reform movements, however, specify the characteristics of an assessment that is likely to affect the quality of education: a focus on student achievements; an assessment carried out by an agency outside the school; and the expectation that assessment will act as a lever of reform. All three characteristics apply to national and international assessments, which will be described in the following chapters.

IV. National assessments of student achievement

National assessment (sometimes called system assessment, learning assessment, and assessment of learning outcomes) is the area of assessment that has received most attention in the context of improving the quality of education since the 1990s. In this chapter, a number of aspects of this form of assessment will be addressed: the nature of national assessment; models of national assessment; reasons for carrying out an assessment; the information needs it meets; and growth in assessment activity (see Greaney and Kellaghan, 1996; Kellaghan, 1997; Murphy, Greaney, Lockheed and Rojas, 1996). Regional or state-wide assessments which are carried out in a number of countries (e.g. Argentina, Australia, Brazil, India, and the United States of America) are considered under the general heading of national assessment, since they also are a form of system assessment and share many characteristics with national assessments.

What is a national assessment?

A national assessment may be defined as an exercise designed to describe the level of achievements, not of individual students, but of a whole education system, or a clearly defined part of it (e.g. fourth-grade pupils or 11-year olds). A number of countries have up to now had experience of only one national assessment. In others, a national assessment system may be said to be in place, as assessments have been carried out for some time on a more or less regular basis.

Similarities and differences are evident in national assessment systems across the world. All include assessment of students' first language and mathematics. Science is included in some, as well as a second language, art, music, and social studies. In practically all systems, students are assessed at the primary-school level. In most systems, students are also assessed at secondary school, usually during the period of compulsory education. Systems differ in the frequency with which assessments are carried out. In some, an assessment is carried out every year, although the curriculum area that is assessed may vary from year to year. In other systems, assessments are less

frequent. Participation by a school may be voluntary, or it may be mandated. When voluntary, non-participation may bias assessment results so that they do not provide an accurate reflection of achievement levels in a system.

Since national assessments are relatively new on the educational scene, many educational planners and administrators may be only vaguely familiar with what they involve. They may be aware that such assessments fit broadly into the category of surveys, and that data are collected in a large number of schools using assessment instruments and questionnaires. However, they may not be aware of the many steps involved in carrying out an assessment.

While assessments vary, many feature the elements set out in *Box 4.1*. The centrepiece of the assessment is the collection of data in schools. Usually students respond to assessment instruments or tests and questionnaires in groups, and the total engagement time for a student is unlikely to be more than a few hours. Teachers may be requested to complete questionnaires in which they provide information considered relevant to an interpretation of their students' achievements. In some cases, they also take the achievement test to which students are responding. The administration of instruments to students may be entrusted to teachers in a school, or examiners may be sent in from outside the school. In some national assessments, in either case, supervisors visit a sample of schools to ensure that procedures are properly followed.

Box 4.1 Main elements of a national assessment

- The Ministry of Education (MOE) appoints an implementing agency either within the Ministry or an independent external body (e.g. a university department or a research organization) and provides funding.
- Policy needs to be addressed in the assessment are determined by the Ministry, sometimes in consultation with key educational stakeholders (e.g. teachers' representatives, curriculum specialists, business people, parents).
- The MOE, or a steering committee nominated by it, identifies the population to be assessed (e.g. fourth-grade students).
- The area of achievement to be assessed is determined.
- The implementing agency prepares achievement tests and supporting questionnaires and administration manuals.
- The tests and supporting documents are pilot-tested, and subsequently reviewed to determine curricular and technical adequacy.
- The implementing agency selects the targeted sample (or population) of schools/students, arranges for printing of materials, and establishes communication with selected schools.
- Test administrators (classroom teachers, school inspectors, or graduate university students) are trained by the implementing agency.
- Survey instruments (tests and questionnaires) are administered in schools on a specified date.
- Survey instruments are collected, returned to the implementing agency, cleaned, and prepared for analysis.

Considerable time will be required before fieldwork for an assessment is carried out in schools to prepare for data collection (e.g. constructing assessment instruments), and, following fieldwork, for data analysis and the preparation of reports. It is important that the student achievements that are assessed are important outcomes of schooling; that the method of sampling ensures that the data that are collected adequately represent the achievements of the education system as a whole (or of a clearly identified part of it); and that analyses identify and describe the main features of the data that have been collected, including relationships between significant variables. All these activities will require considerable resources and political support.

Models of national assessment

There are what might loosely be described as two basic 'models' of national or system assessment, both of which have changed in a number of ways over the years. One is derived from the United States National Assessment of Educational Progress (NAEP) which began in 1969 (see Johnson, 1992); the other from the most recent embodiment in the United Kingdom of assessment which was introduced with the national curriculum in 1988 (see Daugherty, 1997). Both approaches seek to obtain empirical data that can be interpreted as indicative of learning in specific curriculum areas at clearly identified levels of the education system. However, they differ in a number of important respects, and these differences have implications for interpretation, impact, and cost.

The first way in which they differ is that the American model uses a sample of schools/students, while in the UK all students at the relevant age level participate in a census (population) operation. Chile also has a long tradition of census assessments. Second, the American model adopts an analytic view of achievement, and relies on assessment administered under standardized conditions in a few sessions. Assessment in the British model is holistic and performance-based, is integrated into everyday classroom practice, and is administered over several days by teachers, thus avoiding the artificial separation of assessment from teaching and learning (though this aspect has been modified and there has been movement towards the use of more standardized procedures). The New Zealand National Education Monitoring Project (NEMP) also involves a series of complex performance tasks, although in this case the assessments are carried out by pairs of specially trained teachers from other schools (see *Box 4.2*).

Third, in the American national assessment model, each student takes only a fraction of a large number of assessment tasks (matrix sampling). This allows for extensive curriculum coverage without requiring students to spend a lot of time responding to tasks. In keeping with its holistic view of achievement, there is no provision in the British model for systematically allocating different tasks to different

students. Fourth, the American model does not provide information about individual schools or districts; the British model does.

Fifth, an important function of the American system is to monitor achievement over time. While there have been studies of change based on data obtained in the British system, lack of standardization in assessment procedures renders comparisons over time of doubtful validity. Finally, the American system was not designed to impact directly on schools and teachers, although the information derived from an assessment might be used to reach policy decisions that would eventually impact on schools and teachers. The British system, on the other hand, was designed to have such an impact, firstly by providing teachers with exemplars of good assessment practices which, it was hoped, would influence their curriculum priorities and methods of assessment, and secondly through the publication of information on the performance of students in individual schools.

Box 4.2 The New Zealand National Monitoring Programme

In the New Zealand National Monitoring Programme, students at Grades 4 and 8 take a wide range of tasks, requiring up to four hours over the course of a week. The programme differs from the British one in that the tasks are administered by teachers from schools other than the school in which students are being assessed.

In the case of the assessment of social studies, four approaches are used: (i) eleven tasks are administered in one-to-one settings in which students use material and visual information and respond orally; (ii) ten tasks are presented to small groups of students working together; (iii) for five tasks, students work independently and record their responses on paper; (iv) for a further three pencil-and-paper tasks, four students work independently.

Source: Flockton and Crooks, 1998.

Both models are operated in France (see Michel, 1995). In one operation, samples of students are selected about every five years at the end of Grades 7, 9, and 10 to provide information on achievements at the system level in relation to the national curriculum. In a separate operation, the total population of students takes 'diagnostic' assessments on the commencement of study in Grades 3, 6, and 10. The latter assessments are designed to provide information on individual schools, and feedback is sent to schools, students, and students' parents, with the objective of assisting teachers in adapting their pedagogy to the needs of their students.

Although NAEP in the USA still uses samples of students, most states in that country have now adopted aspects of the British model and have put in place programmes to assess the learning of all students, and to hold teachers and/or students accountable for learning outcomes. When assessments allow the identification of individual school districts, schools, teachers, and even students, they may be considered as surrogates for public examinations. The development of this type of assessment has been most marked in countries that do not have a public examination system (the USA and Latin American countries) and countries in which, although they do have such a system, examinations are not held at the grade levels at which the assessments are carried out (United Kingdom, France).

Both models are also used in Brazil (see Crespo, Soares and de Mello e Souza, 2000). As in the USA, a sample of schools and students is selected for the national assessment (*Sistema Nacional de Avaliação do Ensino Básico*). In individual states, the assessment extends to all schools.

Although national or state assessments that are carried out with a total population of students are also a feature of a number of other countries (e.g. Argentina, Chile), their most extensive and consistent elaboration is to be found in the state-mandated assessments of the USA. These assessments are complex and varied, and changed year by year through the 1990s on the basis of perceived effects and feedback from educators, policy-makers, and the public. Not all aspects of programmes are yet in place in all states. Variation exists in

assessment procedures, in the grades and curriculum areas assessed, and in the consequences of performance on the assessment tasks for low-performing individuals and institutions. States also differ in how widely assessment results are disseminated, and in whether or not results are presented for individual school districts and schools.

In general, states are more likely to attach high stakes to performance for teachers, schools, and districts than for students. However, in 22 states, the consequences for both students and teachers/institutions are significant. For students, they can involve non-promotion from grade to grade, or failure to obtain a high-school diploma. For teachers they can mean salary supplements or a decision to terminate an appointment as principal. For schools, they can mean loss of accreditation. In a smaller number of states, sanctions are limited to the public dissemination of results (Shore, Pedulla and Clarke, 2001).

Why carry out a national assessment?

The motivation for carrying out a national assessment is not always clearly articulated. However, an examination of national assessments throughout the world reveals a number of motivations.

- National assessments are carried out with the intention of raising standards. The need to raise standards may arise from a perception that they have deteriorated, or a perception that higher levels of knowledge and skills are required in the workplace if a country is to be economically competitive.

- National assessments are carried out to help maintain standards. In Thailand, for example, a national assessment at the end of secondary education (Grade 12) was introduced because of a fear that the abolition of national external examinations might impact negatively on student achievements. In effect, 'pressure' from the national assessment was to replace 'pressure' from the national examination.

- National assessments are carried out to provide information that can be used to aid decisions about the allocation of resources. Decisions may involve making resources available to:
  - the education system in general (e.g. for the reform of curricula, for teacher development);
  - categories of school with particular characteristics to redress inequities (e.g. schools in rural areas, schools serving students in socio-economically disadvantaged areas); or
  - individual schools (e.g. as a reward to schools in which students perform well on the assessment).

- National assessments are carried out to obtain information that can be used to assign accountability for student performance.

- National assessments are carried out as part of a modernization movement, perhaps under the influence of donor agencies, sometimes with little attention to sustainability or to how the information that is generated will be used.

- National assessments are carried out to alter the balance of control in an education system. An assessment may be used to ensure that what is taught in schools is less dependent on the professional judgement of teachers, and more dependent on central authorities that mandated the assessment. It may also be used to alter the balance of control between a central authority and a local authority. An Egyptian national assessment, for example, was perceived by local authorities to be a means of increasing central control of the education system. By contrast, state assessments in Argentina were inspired in part to signal independence from central control.

- National assessments are carried out to compensate for the poor assessment practices of teachers. In Egypt, the introduction of a national assessment involving all schools was prompted by problems of inefficiency (a high rate of early school leaving and a high rate of grade repetition), which were attributed to deficiencies in teachers' assessment practices. It was believed that teachers lacked clarity about the objectives or procedures of assessing students; that the standards of achievement they expected students to attain were not informed by a knowledge of national or regional

norms; that local examinations were of poor quality; and that assessment practices were not linked to guidelines for remedial action. It was envisaged that the national assessment would provide teachers and parents with higher quality information about students' achievements; would identify students at risk of educational failure; and would provide a basis for remedial action (Carroll, 1996).

Information needs addressed in a national assessment

The information needs addressed in a national assessment, and the procedures that are chosen for an assessment, will depend on such diverse factors as the interests of key stakeholders, the level of policy-making sophistication, views about the nature of education and accountability, and the capacity of the implementing agency. The needs of policy-makers and managers will be paramount. However, the involvement of other stakeholders in the determination of needs can also be beneficial insofar as it ensures that their concerns are represented. Since the educational-political power structures of countries differ, the stakeholder interests that are represented in a national assessment will vary from country to country.

Although the information needs addressed in national assessments vary, all assessments seek answers to one or more of the following questions.

- How well are students learning in the education system (with reference to general expectations, the aims of the curriculum, or preparation for life)?

- Is there evidence of particular strengths and weaknesses in students' knowledge and skills?

- Do particular sub-groups in the population perform poorly? Are there, for example, disparities between the achievements of boys and girls, of students in urban and rural locations, of students from different language or ethnic groups, of students in different regions of the country?

- What factors are associated with student achievement? To what extent does achievement vary with characteristics of the learning environment (e.g. school resources, teacher preparation and competence, type of school) or with students' home and community circumstances?

- Do the achievements of students change over time? This question may be of particular interest if reforms of the education system are being undertaken. To answer the question, it will be necessary to carry out assessments that yield comparable data at different points in time.

Growth in national assessment activity

While most industrialized countries have had systems of national assessment for some time, some of which are established on a statutory basis (France, United Kingdom), it is only during the last decade that serious efforts have been made to develop the capacity to administer national assessments in other parts of the world. Most countries have followed the American sample-based approach and have used standardized assessment instruments. However, a number have carried out assessments in all schools.

The UNESCO-UNICEF Monitoring Project, in conjunction with the Education For All initiative, has been responsible for initiating in close to 50 countries sample-based assessments focused on basic learning competences in literacy, numeracy, and skills required for daily life (relating to awareness and knowledge of health, nutrition, sanitation, hygiene) after four years of schooling (Chinapah, 1997). The aim of the project was to provide decision-makers with up-to-date data that would allow them to measure progress, and to identify the most serious shortfalls in attaining the objectives of basic education. Twenty-four of the participating countries are in Africa, twelve in Asia, five in the Middle East, three in Latin America and the Caribbean, and three in Europe (Kellaghan and Greaney, 2001).

During the 1990s, a number of ministries of education in southern Africa have been working in association with the International Institute

for Educational Planning (IIEP) to provide training for educational planners in the technical skills required to monitor and evaluate basic education systems. A proposal prepared by a group of educational planners to create an association known as the Southern Africa Consortium for Monitoring Educational Quality (SACMEQ) received a positive reaction from ministries of education and was officially launched in 1995. Between 1995 and 1998, seven education ministries collected information in SACMEQ I on baseline indicators for educational inputs, general conditions of schooling, equity assessments for human and material resource allocations, and the literacy levels of Grade 6 students. Fifteen ministries participated in SACMEQ II between 1998 and 2001 (Ross et al., 2000).

Viet Nam's Ministry of Education and Training called upon IIEP's expertise in sampling, file merging and cleaning, for its Grade 5 survey of achievement in reading and mathematics. There were 3,660 schools in the sample, from each of which 20 students and two Grade 5 teachers were drawn at random.

While SACMEQ is considered here under national assessments, it also has an international dimension, since research questions, instruments, sampling procedures, target populations, and analyses are common to all participating countries. Thus, although comparative cross-national data are not presented in the seven national SACMEQ I reports that have been prepared (e.g. Machingaidze, Pfukani and Shumba, 1998), comparisons across countries are possible. Policies on presenting comparative data are set down by the participating countries, and in the SACMEQ I study, it was agreed that only national, and not cross-national, reports would be published. There are some indications that this may be reviewed for the SACMEQ II study.

There was also a rapid development in the establishment of national assessments during the 1990s in Latin American and Caribbean countries. The assessments were often associated with the provision of baseline data for educational reforms. Practically all ministries of education in the region have now incorporated national assessments into their agenda (Rojas and Esquivel, 1998). There is considerable similarity, but not uniformity, in the curriculum areas assessed.

Language/communication and mathematics figure in all; and science and social studies or history and geography in about half. Most assessments were carried out in primary-school grades. Both census and sample procedures have been used.

Apart from national assessments supported by UNESCO-UNICEF and IIEP, several countries have carried out assessments with the support of the World Bank and other agencies (e.g. Colombia, India, Namibia, Nepal, Uganda, Zambia) (see, e.g. Kelly and Kanyika, 2000).

Conclusion

National assessments represent a relatively new form of activity designed to provide information about the level of achievements of an education system, or of a clearly defined part of it. While they differ from more traditional forms of assessment which focused on individual students, they draw on such assessments for guidance in constructing assessment instruments and, in fact, estimates of system achievements are based on an aggregation of individual student performances. However, there are a number of issues that arise in a national assessment that do not arise when an assessment is being carried out to make a decision about an individual student. Some are technical, and relate to the construction of assessment instruments, sampling, and data analysis, and some are political, relating to the determination of policy needs and obtaining the support of key stakeholders.

Ideally, a decision to participate in a national assessment should follow an analysis of the information needs of policy-makers, educational planners, and administrators, who may be assisted in this task by other stakeholders in the education system. While enlisting the involvement of key stakeholders can help build support for an assessment, some problems and some degree of conflict and resistance can be anticipated. A national assessment in Egypt led to conflict between central government and local authorities, while in England, teachers on one occasion boycotted the administration of assessments.

Information in most national assessments is sought about the learning achievements of students, about strengths and weaknesses in

students' knowledge and skills, about the performance of sub-groups in the population (e.g. by region, by gender), and about correlates of achievement. However, there is considerable variation between assessments in the curriculum areas that are assessed; the grade or age level at which an assessment is carried out; elements of the population that are excluded (e.g. ones that are too difficult to reach or that would not be able to respond to the assessment instrument); the nature of the assessment instrument or process; whether all students at a particular age or grade level or a sample of students are assessed; how results are presented; how the information obtained is communicated and used; and who is regarded as accountable for student performance (see Ross and Mählck, 1990).

The novelty and complexity of a national assessment should serve as a reminder of the need to ensure that adequate personal and financial resources are available to carry out the task in a competent and timely fashion. Unfortunately, many national assessments fail to meet required technical standards relating to instrument design, sampling, test administration, data preparation, data analysis, or data interpretation. As a consequence, the data obtained may provide inaccurate information, making its use for immediate policy purposes or for monitoring over time problematic.

V. Describing and using information derived from a national assessment

It is important that the information derived from a national assessment be presented in a way that will be meaningful to policy-makers, educational planners and managers, school personnel, and others for whom it may have implications. In the first part of this chapter a variety of procedures that have been used to report performance are described. Although some assessments include measures of non-cognitive characteristics of students (e.g. students' attitudes to, and interest in, a curriculum area), the procedures described here are limited to measures of students' cognitive achievements. Some present a detailed analysis of students' achievements on an assessment task. Others are criterion-referenced, examining actual performance in the context of expected performance, or describing performance in terms of levels of proficiency. Others examine student performance in a comparative context, while some procedures include the results of analyses in which relationships between achievement and other aspects of an education system are described. Each approach has advantages and disadvantages, and may be more appropriate to the needs of some users than to the needs of other users.

The communication and use of the findings of a national assessment are discussed in the second and third sections of the chapter.

Describing student achievements

The presentation of results for a national assessment differs from their presentation in the case of more traditional forms of assessment in a number of ways. The main difference arises from the fact that the primary interest in a national assessment is in the aggregated performance of students, not in the performance of individual students, as is the case, for example, in external (public) examinations. Thus, a description of performance in a national assessment may be regarded as an *indicator*, which is a quantifiable measure used to describe and monitor the education process. Indicators should help increase

understanding of how the process is working, and may provide a basis for diagnostic and accountability procedures (see Burnstein, Oakes and Guiton, 1992).

A basic question to be addressed as far as an indicator representing student achievement is considered is: how is the performance of a group of students to be represented in a meaningful way? Several approaches have been adopted. In what is perhaps the simplest of these, the mean of raw scores or the percentage of items answered correctly by students is calculated. More sophisticated methods involve the scaling of student responses. One such method, which is being used with increasing frequency, involves modelling data on item response theory (IRT). It has been used to address problems arising from the design of assessment instruments, and when there is a need to equate performances on two or more instruments. It was used, for example, to place the results of the American national assessment of reading obtained in different years on a common scale. Some controversy surrounds IRT, in particular its assumption that a single underlying ability or trait adequately represents the complexity of achievement.

Reporting the results of an assessment in terms of an average of raw, percentage, or scaled scores does not, in itself, convey a great deal of information. In recognition of this limitation, a number of other strategies have been adopted in analyzing and reporting data. These may be categorized under four headings: analysis of student achievements; norm-referenced comparisons; criterion-referenced comparisons; and examination of relationships between achievement and other indicators.

Analysis of student achievements

Performance on individual items. However results of an assessment are summarily presented, a more detailed description of students' performance may reveal information regarding weaknesses in teaching and learning. For example, weaknesses may become apparent in a display of the percentage of students who answer each item correctly on an assessment task. While this information may be too detailed for most individuals, for those prepared to take the time

or with a special interest (e.g. curriculum specialists), a perusal of it can provide insights into the areas of the curriculum in which students are achieving, and areas in which they are not.

Performance in curriculum domains. An alternative to reporting results in such detail as percentages responding correctly to individual items in an assessment, is to stratify the curriculum into a number of domains, and to report students' scores for each domain. For example, performance in mathematics might be described in terms of student mean scores in the content areas of whole numbers; fractions and proportionality; measurement; estimation and number sense; data representation; and patterns, relations, and functions. Performance in reading may be described in terms of student mean scores on items that present text of different types: narrative prose (continuous text in which the writer's main aim is to tell a story – fact or fiction); expository prose (continuous text designed to describe or explain factual information or opinion); and documents (structured information displays presented in the form of charts, tables, maps, or notices). Many national assessments provide information on the performance of students by domain. However, in the absence of further analysis, this information will not reveal the precise knowledge/skills that students have acquired in each domain.

Performance defined in terms of cognitive processes. Some assessment instruments are designed to provide evidence of students' use of hierarchically defined cognitive processes in assessment tasks. For example, in the case of reading, levels of text processing may be categorized in ascending order of complexity as: literal response (or verbatim matching); paraphrasing; identifying the main idea in a text; making inferences from the information provided; and locating and processing information. In a national assessment in Palestine, sixth-grade students were found to have difficulty in identifying the main ideas in texts, and in dealing with tasks that required higher levels of comprehension (Palestine Ministry of Education, 1998). *Box 5.1* contains conclusions from an assessment of reading in Ireland regarding problems revealed by fifth-grade students in the interpretation of text, as well as in their ability to deal with different kinds of material.

Box 5.1 Student problems in reading revealed in a national assessment

In general, pupils performed better in the Narrative and Documents domains on Local questions and questions which require fewer steps for their solution than on Textplus questions and questions which require higher levels of cognitive complexity and more steps in their solution. Analysis of pupil responses points to the following conclusions. Pupils had difficulty in differentiating between the passive and the active voice; identifying the theme of a text and differentiating this from the protagonist or cause of action; identifying the motives for an action and differentiating these from the outcomes of an action; identifying and understanding causality; inferring the meaning of a specific word from the text using context clues; understanding the concept of function and inferring function from the text; using indexes, tables of contents, and dictionaries; cross-referencing on a timetable or graph across two or more dimensions; and integrating distinct pieces of information from continuous texts, such as story texts, and non-continuous texts, such as maps and charts, to produce an answer which is not apparent in the stimulus.

Source: Cosgrove, Kellaghan, Forde and Morgan. 2000, pp. 43-44.

Norm-referenced comparisons

A norm-referenced comparison indicates where a group stands relative to another group or other groups that responded to an assessment instrument.

Performance of a comparable group. The mean score on a national assessment may be compared with the mean score of students in an earlier assessment. The comparison may be made in the context of an educational reform programme; the expectation would be that mean level of performance would increase following a reform or that, at least, it would not decrease in a reform that increased the proportion of children at a particular age attending school.

The possibility of monitoring the achievements of an education system over time is an attractive feature of a national assessment system. Three difficulties associated with this, however, should be

recognized. First, if different assessment instruments are used, a common scale that links performance on the instruments has to be constructed. However, in this situation, it is very difficult to ensure that observed differences in performance are not due to differences in the difficulty of instruments. Second, if change is being monitored over a period of time in which changes to a curriculum are being implemented, an assessment instrument that was appropriate before the curriculum change might not be appropriate after it. This will be the case if the revised curriculum differs from the curriculum that it replaces in its objectives, content coverage, or recommended implementation procedures. Third, the interpretation of change over time, or lack of it, will have to take into account any changes in participation rates in education that have occurred over the period for which comparisons are being made. Thus, for example, in interpreting the achievements of students in the education system following the introduction of universal primary education in Uganda, it would be necessary to have in mind the implications of ways in which the characteristics of students in the system may have changed as a result of higher participation rates in disadvantaged rural areas, as well as the impact on schools of having to deal with larger numbers of students without a commensurate increase in resources.

Performance of sub-groups in an assessment. Comparisons may be made of the performance of sub-groups in an assessment, as well as with the performance of other groups. While evidence has long been available in many countries relating to access to, and participation in, education of a variety of sub-groups, national assessment data can bring this a step further by providing evidence about the achievements of the sub-groups. If sample sizes are sufficiently large, data can provide evidence on achievement by gender, region, sector (urban/rural), ethnic or language group membership, type of institution attended by students (public/private), or indeed any other variable that might be of interest.

Criterion-referenced comparisons

A criterion-referenced comparison involves the interpretation of scores, not with reference to the performance of other students (which is the case in norm-referenced comparisons), but with reference to specific knowledge and skills, which often are based on objectives

specified in a curriculum. Criterion-referenced comparisons have been used widely to place individuals above or below specific performance standards such as those associated with 'minimal competency' or 'mastery'. In national assessments, three forms of criterion-referenced comparisons have been used.

Mastery of curriculum objectives. Student performance in some assessments is scored in terms of whether or not students have achieved 'mastery' on a series of curriculum objectives. A student will be assumed to have achieved mastery if he or she answers correctly a predetermined number of the items that assess an objective or, in more technical approaches, on the basis of the probability of answering correctly items at the level of mastery. The results of the assessment will be presented in terms of the proportion of students that master each objective. Some curriculum areas (e.g. mathematics, science) lend themselves more readily to this approach than others. Examples of objectives in mathematics for students at the end of primary schooling might include 'student can complete simple number sentences'; 'student can identify the distributive property'; 'student can identify the least common multiple of two numbers.'

Since a large number of objectives are likely to be included in a curriculum specification, a national assessment that describes performance in terms of mastery of objectives will yield an equally large number of scores. A positive feature of this is that the information might provide evidence of differential success in mastery of objectives, which, in turn, could form the basis for further exploration of problems in curriculum implementation or for the design of in-service teacher-training programmes. A negative feature is that policy-makers and the public, as in the case of information on performance on individual items, may find the large amount of data that is generated difficult to absorb and interpret.

Performance on attainment targets. A further possibility in describing performance on a national assessment is to indicate the proportion of students achieving attainment targets in the curriculum. This can be done if a curriculum is structured in terms of the level of performance expected of students at particular ages. For example, the national curriculum in England and Wales defines eight levels up to

age 14 in the main school subjects. National assessment reports indicate the proportion of students who reach each level of performance.

Performance at varying levels of 'performance standards'. In some assessments, the continuum of student achievements is divided into two or more levels (e.g. 'satisfactory', 'unsatisfactory'; 'basic', 'proficient', 'advanced') and the proportion achieving at each level is reported. This is obviously a very attractive way of presenting results, not just for educational planners and administrators, but also for the general public. It would seem particularly useful if national assessments are carried out regularly, as it would appear to provide a readily understood categorization for comparing standards at different points in time.

The attractiveness of categorizing national assessment data in this way has led in some countries to the adoption of criteria used in other assessment systems to represent performance. Thus, students who obtain a score of 40 per cent (the pass mark on an external examination) might be classified as having 'passed' or as being 'proficient.' This approach is largely arbitrary since it is not based on the analysis of student performance that would be necessary before judging it 'proficient'.

While several more sophisticated procedures have been devised to identify the point or points on the continuum of achievement at which it can be said an individual meets or satisfies a standard, all also are arbitrary to some extent. As noted in *Chapter II*, most procedures are based on judgements about how students of varying levels of 'competency' would perform on the assessment, and so are subject to human error. It is not surprising that different methods result in quite different proportions of students being judged as having reached a standard.

Efforts to describe the knowledge and skills that represent performance at each level of proficiency ('achievement-level descriptions') have been important, both for the production of valid standards and for the communication of information about the meaning of standards. In these procedures, panels of judges, with representation of curriculum specialists, the public, parents, policy-makers, the business

community, and teachers, map scores on an assessment task on to performance levels; that is, particular types of knowledge and skills are matched with scores on the task to provide a picture of what students classified at varying levels of proficiency know and can do (Baker and Lynn, 1995). Such efforts have proved more difficult than was originally anticipated, and have given rise to considerable controversy.

Relationships between achievement and other indicators

In many national assessments, data are collected in questionnaires from students, teachers, headteachers, and sometimes parents, about such factors as characteristics of students' home background and community, student attitudes and aspirations, aspects of teaching and learning (such as the amount of time devoted to teaching a particular topic), and school resources and organization. This information is of interest in its own right, and can reveal much about the quality of educational provision. It has been used to describe the context in which schools operate (e.g. by taking into account the socio-economic characteristics of the student body) when comparisons are made between schools, and to reach decisions about the allocation of additional resources to schools (see *Chapter VII*).

Background data may also be collected with the intention of correlating them with the achievements of students in the expectation that the results of analyses will throw light on the school and student background factors that contribute to achievement, and so merit consideration in educational policy-making and decision-making. However, appealing though this prospect may be, the usefulness of information derived from an examination of relationships between students' achievements and other information obtained in a national assessment is limited by a number of factors. Background data provided by teachers and students in questionnaires may not be accurate, while personal characteristics such as attitudes, values, and aspirations are extremely difficult to measure in a reliable way. Furthermore, the actual analysis of data poses serious challenges. It is difficult to disentangle the unique contribution to student achievement of factors that are themselves interrelated, as it is to distinguish between the effects of community, home, school-level, and individual student-

level factors. The fact that analyses are based on cross-sectional data also limits the inferences that can be made about factors that are found to be associated with achievement. Association does not necessarily imply a causal relationship. Inferences regarding the 'causes' of achievement can only be made if supporting evidence is available from other sources.

A consideration of these limitations should not lead to the conclusion that there is no value in collecting background information in a national assessment. Such information can be of assistance in the interpretation of findings and may, for example, be useful in identifying types of schools which, following further investigation, may require additional resources.

The communication of information derived from a national assessment

One of the arguments made in support of national assessments is that indicators describing achievements can provide information that is broadly useful to a variety of publics with diverse needs and interests. Thus it is appropriate in planning an assessment to consider how the information it will yield will be fitted into the managerial, administrative, and accountability structures of the education system, and how it will be communicated and used. There would seem to be considerable variation in the extent to which this happens. In the SACMEQ studies, ministries of education were involved from the beginning in setting the questions that the studies were designed to answer. A great deal of effort was also put into ensuring that final policy suggestions were agreed within the top ranks of the ministry before publication. In other studies, the issue of how information might be used received little attention. There are cases in which data have not been made available to educational planners and managers, while in at least one country collected data have not been analyzed.

Assessment data should be of particular interest to educational policy-makers and administrators, such as senior ministry of education officials (planners and managers) and politicians. Without the more 'objective' information that such data convey, policy-making can be

unduly influenced by the personal biases of individuals, vested interests of school managers or teacher unions, and anecdotal evidence offered by business interests, journalists, and others. This is not to say that decisions will ever be made solely on the basis of assessment information, even when available. Value considerations and the pressures of powerful constituencies will always have their say. However, the availability of information from an assessment should serve to make decision-making more reflective, and to strengthen its factual basis.

Policy-makers, planners, and managers should also ensure that an assessment is in accord with other instructional guidance mechanisms in the education system. If, for example, the skills and knowledge assessed in a national assessment differ from those assessed in a public examination to which high stakes are attached, teachers will be unlikely to pay much attention to the assessment results in their teaching. To address this situation, it is necessary to ensure that the knowledge and skills that are measured in an assessment and in an examination, and so may be regarded as important for students to acquire, are compatible. Other components of the education process that convey policy guidance (e.g. textbooks, instructional materials, and pre-service and in-service teacher education) should also be aligned with, and reinforce, the assessment.

If, as is likely, the findings of an assessment indicate the need for new approaches to teaching and learning, teachers throughout the education system will need to be provided with relevant information. When data are available about all schools, the publication of school results in league tables might be interpreted as signalling to some schools the need for improvement. This in itself might prompt action, but even in this case, the need for more information about deficiencies, and how these might be addressed, is usually recognized.

There are a number of ways in which a national assessment can impact on teachers. In some assessments, teachers are involved in the development of instruments and/or in the scoring of student responses. In others, they carry out the fieldwork for an assessment. The experience of performing these tasks should help improve teachers'

assessment skills, as well as develop their understanding of the reforms that underlie an assessment. However, more formal and more focused approaches to reach all teachers will also be required.

There are several examples of efforts to communicate national assessment results to teachers. Following an assessment in Colombia, results were released through the mass media, and a national seminar was held to identify appropriate strategies for improving educational quality. This was followed by national and local workshops to discuss results and their implications. Information on national and regional results was sent to individual teachers in newsletters and brochures. 'Profiles' of high-scoring schools were prepared containing information on such factors as the number of hours in a week devoted to a curriculum area; teachers' emphases on specific content areas; teachers' educational level; school facilities; and the number of textbooks available per student.

In Argentina, the authority responsible for the national assessment produced a series of publications for teachers, principals, and school supervisors. Specific deficiencies in schools were identified, and recommendations for addressing them were incorporated into pre-service and in-service teacher-training courses. A video series for teachers was also produced. In Uruguay, confidential item-by-item reports on students' performance on an assessment were sent to schools to provide a basis for the diagnosis of children's difficulties.

In Uganda, dissemination focused on assisting district educational authorities in developing strategies to apply, in the classroom, information derived from national assessments carried out by the Uganda National Examinations Board. The process comprised a series of workshops and seminars for teacher trainers, school inspectors, and teachers in which the results of an assessment were presented. Implications for teaching were considered, as well as how to use information to adapt the instructional process to improve learning. Teachers were expected to shift their emphasis in the way they presented curriculum material, and to pay more attention to areas of knowledge and skill that the national assessment identified as being relatively weak. It was decided to involve all teachers in schools, not

just teachers at the grade level targeted in the assessment, first, because the roots of problems that might be identified were likely to be found in classes other than the one in which students were assessed, and, secondly, it was considered desirable that problems should be addressed in a consistent way throughout a school. The need for continuing support to teachers as they attempted to implement change was recognized.

While such efforts are to be commended, the difficulty of conveying what is meant by centrally specified standards and information derived from assessments to those who have to act on the information should not be underestimated. In an evaluation of the national assessment in Brazil, it was found that reports were not written in a user-friendly language, and that only a few people at state level had the critical capacity to interpret results. It would be surprising if this was not also the case in other countries.

Difficulties in communicating with teachers are likely to be even greater. The gulf between what is required in reforms and what teachers understand and actually do can be very large. It cannot be assumed that teachers will know what is really involved in reform proposals, or what they should do on the basis of assessment results. Even when teachers are provided with in-service training and guidance, the meanings attributed to reform by policy-makers and by teachers may be disparate, since guidance is likely to be interpreted in the context of teachers' existing understandings and practice relating to the curriculum and instruction, which may be very different from those of the reformers.

The use of national assessment data

System level

Not a great deal of information is available about the use of the findings of national assessments even when these are communicated to those involved in the educational process. This does not mean that findings are not used, but that use is not often publicly documented. At the same time, given that national assessment activity is relatively

new in many countries, it would not be surprising to find that its potential value is not fully appreciated, or that procedures to act on findings are not fully developed. Indeed, it took some considerable time for the findings of the National Assessment of Educational Progress (NAEP) in the USA to gain much attention, and much longer to impact in any way on educational policy or practice.

There is some evidence that the findings of national assessments have been used at the system level in a number of countries. One use has involved curriculum revision. For example, in Thailand, the poor performance of students in areas of mathematics and science led to greater curricular emphasis on process skills and the application of knowledge. In Brazil, the findings of national assessments have also been associated with curricular changes and pedagogical innovations, and have influenced human resource training and policies regarding the financing of schools serving disadvantaged areas. Another example of the use of national assessment information comes from Chile, where information was used to decide how supervisory staff are used.

School level

Uses of the findings of national assessments to evaluate schools are described in *Chapter VII*. At this stage, a number of general points will be made, in particular regarding inferences about accountability that can be made from use.

It may be assumed that responsibility for addressing any deficiencies that may be identified in a national assessment lies with the institution or person that is considered accountable (see *Chapter II*). The data collection process may provide some clues as to who this is. When data are collected in a sample of schools, participating schools are usually assured that results for individual schools will not be identified. Anyhow, only a relatively small (though representative) number of schools will take part in the assessment. Thus, it will not be possible to attach sanctions to the performance of individual schools, teachers, or students throughout the country. In this case, it may be assumed, in the absence of evidence to the contrary, that schools will not be held primarily responsible for student performance.

Assessments in which information is obtained on all schools (e.g. in the United Kingdom, France, Chile, Uruguay), on the other hand, have the capability of identifying poorly performing schools and teachers, and it is not a very large step from this to allocating responsibility for student performance primarily to them. The inference that the school is being held accountable certainly seems warranted when assessment results are published in league tables.

Collecting assessment data in all schools, however, does not necessarily imply that responsibility for performance will be laid at the door of the school or teacher. Although data are collected in all schools in Uruguay, central government holds itself accountable for educational quality and for intervening on behalf of impoverished communities, rather than placing the responsibility on individual schools. Indeed, the focus of its national assessment has been to identify and address the social needs that condition student learning. Results for individual schools or teachers are not published, and assessment data are used primarily for diagnostic purposes (Benveniste, 2000a).

A similar situation exists in Australia's Capital Territory, which decided against using data from its programme of assessment of literacy and numeracy to assess teacher performance, to make positive or negative comments about school performance, or to compare government and non-government schools. Data are, however, used to compare overall results with national benchmarks, to provide schools with confidential information about their performance, to give teachers detailed results on student performance, and to provide parents with data for comparing their child's level of performance with curriculum standards. Schools are also required to prepare reports on how they plan to improve performance levels.

Conclusion

A variety of procedures have been used to describe student achievements in national assessments, some of which describe performance in very detailed terms (e.g. the percentage of students who successfully responded to every item on an assessment task).

While this may have merit for use in an analysis of the strengths and weaknesses of student achievements, it has the disadvantage that it may fail to describe important underlying abilities or understandings, as well as being overwhelming in the detail it provides. Individuals whose main concern is not curriculum analysis will prefer more summary data.

A procedure that has wide appeal is the representation of the performance of students in terms of the level of 'proficiency' they exhibit on an assessment task. In response to this, performance has been described in some assessments, with varying degrees of justification, in terms of the proportion of students who have achieved varying levels of 'proficiency'. In some cases, the specification of levels was arbitrary. In others, serious efforts were made to determine cut-off points between levels and to describe the knowledge and skills that the levels represent. The technical adequacy of even these efforts, however, has been the subject of debate and controversy.

There is no point in carrying out a national assessment if its findings are not communicated to potential users, and are not acted upon. In some cases, little effort seems to have been invested in these tasks; in other cases, strategies have been developed to communicate findings to varying levels in the education system. These recognize that the primary users of national assessment data will be educational policy-makers and administrators, who will be responsible for making decisions about such matters as the adequacy of curricula or the allocation of resources and for integrating an assessment into other instructional guidance mechanisms. Teachers will also need to be involved if proposed action requires them to change the knowledge and skills they teach, or how they teach. Strategies to change teacher behaviour will vary depending on whether a national assessment provides information on all schools in a system, or on only a sample of schools selected to represent the performance of the system in general. In either case, problems associated with changing the behaviour of teachers should be recognized.

VI. International assessments of achievement

In this chapter, some of the characteristics of international assessments are described, as well as examples of how the information provided by assessments has been used. International assessments share many procedural features with national assessments, although they differ from them in a number of respects, most obviously in the fact that they have to be designed to allow administration in more than one country (Beaton, Postlethwaite, Ross, Spearritt and Wolf, 1999; Goldstein, 1996; Greaney and Kellaghan, 1996).

What is an international assessment?

International comparative assessments grew out of a consciousness of the lack of internationally valid standards of achievement with which individual countries could compare the performance of their students. As well as providing data for comparisons, it was envisaged that the studies would capitalize on the variability that exists across education systems, exploiting the conditions which 'one big educational laboratory' of varying school structures and curricula provides, not only to describe existing conditions, but to suggest what might be educationally possible (Husén and Postlethwaite, 1996). Thus, the studies were promoted as having a research perspective as well as more immediate policy implications.

As in national assessments, instruments are developed in international assessments to assess students' knowledge and skills. However, instead of representing the curriculum of only one education system, the instruments have to be considered appropriate for use in all participating systems. The age or grade at which the instruments are to be administered has to be agreed, as have procedures for selecting schools/students. International studies have all been based on samples of students.

Most international assessments have been carried out under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). Since 1961, when a pilot study was

carried out, over 60 countries have participated in studies of achievement in reading, mathematics, science, writing, literature, foreign languages, civic education, and computer literacy. Participation has been more a feature of industrialized countries than of less developed ones. There are several possible reasons for this, including limitations in finances, infrastructure, and human resources, as well as the realization that differences in the conditions of education systems render documentation of associated differences in achievement superfluous and possibly embarrassing.

In 2000, after a number of years of preparation, the Organisation for Economic Co-operation and Development (OECD) launched a Programme for International Student Assessment (PISA), which is designed to collect data at regular intervals on the preparedness for adult life of 15-year olds in OECD member countries and in a number of other countries. Assessment tasks in reading literacy, mathematical literacy, and scientific literacy will be administered every three years (32 countries took part in 2000). A significant feature of PISA is that assessment tasks are based on a 'dynamic' model of lifelong learning in which new knowledge and skills necessary for adaptation to a changing world are continuously acquired throughout life. Thus, as well as assessing the knowledge of 15-year olds, PISA also examines their ability to reflect on, and to apply, their knowledge and experience to real-world issues (OECD, 2001).

Why participate in an international assessment?

Before making a commitment to participate in an international assessment, its advantages vis-à-vis a national assessment merit consideration. A unique feature of international assessments is that they provide an indication of where students' achievements in a country stand relative to the achievements of students in other countries, particularly in countries that may be regarded as economic competitors. This is likely to be the aspect of the studies that attracts most media attention, and may have ramifications for attracting external investment to a country. However much a country's ranking may attract interest, the information that more analytic approaches to examining performance can provide is likely to be much more valuable. Any of

the variety of methods used to describe achievement in a national assessment (described in *Chapter V*) may be used, depending on the nature of the assessment instrument.

International studies also allow an examination of such diverse issues as the significance of age of starting school (which varies from country to country) or the distribution of achievement in schools throughout the country, compared to the distribution in other countries, and the possible reasons for differences (Beaton et al., 1999). It should be recognized, however, that even if a country finds that it is out of step with practice elsewhere on age of school entry, or that the distribution of achievements in its schools is atypical, there might be very little that can be done to change the situation.

A further perceived advantage of an international assessment (which may be a mixed blessing) is that it tends to attract a lot of political and media attention. This can be useful if the publicity translates into greater appreciation among the public of the value of education, and results in increased support for educational expenditure.

Participation in an international assessment also introduces to countries the experience of rigorous sampling, item review, supervision, analysis, and report writing. This can be of considerable benefit to a country in which traditions of empirical educational research and associated technologies (particularly involving survey methodology and the construction of assessment instruments) are not strong.

Issues in international assessments

A number of problems that have been identified in international assessments also merit consideration before deciding on participation (Kellaghan and Grisay, 1995). One relates to the difficulty in designing an assessment procedure that will adequately measure the outcomes of a variety of curricula, given that education systems differ in their goals and in the emphasis and time they accord the domains of a curriculum. There can, however, be a positive aspect to this. If analyses reveal differences between countries in achievement, which in turn reflect differences in curricula (e.g. in content or in the sequence in

which topics are presented), policy-makers might be encouraged to reflect on the appropriateness of the curricula offered in their schools. Is the difference due to choice, because the curriculum suits local conditions, or is it, perhaps, because the curriculum has not kept up with developments elsewhere?

A second problem is likely to arise in international assessments (and may also arise in some national assessments) if it is necessary to translate assessment instruments into one or more languages. If comparisons are to be made between performances assessed in different languages, it should be realized that differences that may emerge may be attributable to language-related differences in the difficulty of assessment tasks. In practice, it is very difficult to ensure that the ways questions are phrased and the cultural appropriateness of content are equivalent in all language versions of an assessment task.

A third problem in international assessments relates to the equivalence across countries of the populations and samples of students that are being compared. For example, differences in performance might arise because countries differ in the extent to which they retain students in their education systems, or in the extent to which categories of students are removed from the mainstream, and so may be excluded from an assessment (e.g. students in special programmes, students in schools in which the language of instruction differs from the language of the assessment).

A fourth problem arises when the primary focus in reporting the results of an international assessment is on the ranking of countries in terms of the average score of their students. Apart from the fact that this might lead to the neglect of more interesting aspects of performance, such as the distribution of achievement within a country, or gender or ethnic-group differences, it also fails to recognize that the differences reflected in ranks might not be indicative of differences that would be considered 'significant' if statistical tests had been applied. This issue has been addressed (though often ignored in the media) in reports of international studies in two ways. In the first, reports indicate whether or not a country mean differs (statistically) significantly from the average score of all participating countries. In the second, the significance of differences between all pairs of countries is determined.

Box 6.1 South African experience with TIMSS and TIMSS-R

Participation in TIMSS and TIMSS-R allowed South Africa to compare its mathematics and science curricula and achievement standards with international standards. It fostered technical capacity in educational assessment, resulted in three reports for policy-makers and presentations to each of nine provincial departments of education, and provided baseline information on achievement standards and a stimulus for additional research.

Much of the time for TIMSS was spent problem solving due to inadequate data on school populations by province, inexperienced management and administrative personnel, and a tradition of working 'normal hours'. For TIMSS-R, more attention was devoted to the content of project manuals supplied by the international study centre, a filing system was installed, quality-control procedures implemented, and external support obtained for sampling (which discovered 4,000 additional schools in the school population). At the logistical level, difficulty was encountered in locating named schools, and it was necessary to employ an external field agency to transport tests (due to an inadequate postal system) and to administer them. Problems encountered during administration included student absenteeism or late arrival, some cheating in overcrowded classes, nervous test takers who believed the test would contribute to their public examination scores, and reluctance to ask for assistance when unsure of test instructions. Many students struggled with open-ended questions, which was not surprising given that roughly 80 per cent took the tests in their second, third, or even fourth language.

Source: Howie, 2001.

A number of other factors are worth considering before making a commitment to participate in an international assessment. Some are political. A country may not wish to take part if poor performance, which may well lead to embarrassment, seems likely. Some countries have actually collected data for an assessment and then, when they saw the results, refused to allow the data to be included in between-country comparisons. Other factors relate to limitations in finances, infrastructure, and human resources. The demands of meeting deadlines may prove very difficult in countries that lack administrative personnel and have to cope with a poor communications infrastructure (see *Box 6.1*).

Obtaining comparative data for neighbouring countries would seem more appropriate than obtaining data for countries across the world that differ greatly in their level of socio-economic development. For this reason, 10 Latin American and Caribbean countries jointly carried out an assessment of basic competences in language and mathematics in 1997. Assessments in southern and eastern Africa by a network of ministries in the 1990s also allow international comparisons at a regional level.

The uses of international assessment data

There are a number of accounts of the use of findings from international studies to inform policy and decision-making. Some describe general reactions to results. Thus, an Italian minister for education is reported to have said that the results of the IEA Six Subject Survey acted like 'an electric shock' to the education system. In the USA and Canada, low scores of students in international mathematics and science assessments influenced public perception of the quality of education provided in these subjects, resulting in the subjects being accorded priority by government.

Specific curriculum changes have also been attributed to the results of international studies (Kellaghan, 1996b). In Japan, it was concluded that the performance in mathematics of students in early studies, although generally good, was relatively better in computation than in application and analysis, which presumably require higher mental processes. As a result, the curriculum was changed to reduce the emphasis on computational skills, a change that was associated with improved performance in application and analysis in a later survey. At a later date, concern over the negative attitudes of Japanese students to mathematics led authorities to propose less rigorous curricula.

On the basis of observed differences in intended curricula between Ireland and other countries in the Second International Mathematics Study (SIMS), alterations were made in the emphasis accorded certain topics in the mathematics curriculum in Irish schools. Commitments to curriculum changes have also been attributed to the Third International Mathematics and Science Study (TIMSS) in Cyprus and

South Africa. In New Zealand, in the preparation of new syllabuses and teachers' guides, special emphasis was placed on strengthening areas of mathematics in which results indicated that the performance of students was relatively poor. The results of IEA studies were interpreted in Hungary as underlining the importance of the home in children's scholastic progress.

The most extensive effort to obtain evidence on the impact of an international study is to be found in the case of TIMSS (Robitaille, Beaton and Plomp, 2000). Twenty-nine education systems provided information on the impact of the findings of the study at the national level, from which a number of general conclusions were drawn. First, in most cases, the results were taken to mean that improvement was called for. Even in some of the highest performing countries, concerns were expressed about aspects of findings (e.g. students' attitudes, or performance on items requiring higher-order thinking skills). Secondly, the results seem to have had a greater impact on the general public and in political circles through the mass media than on educators. Thirdly, TIMSS brought home to policy-makers the fact that educational practices and curricula in mathematics and science were different in other countries, a realization that might lead to curriculum revision.

Conclusion

The main advantage of international assessments over national assessments is that they provide between-country comparative data on student achievement, and most likely also on the conditions of education in participating countries. Furthermore, sharing in the planning and implementation of an international study will reduce costs and provide expertise relating to assessment from outside the country.

International studies, however, have disadvantages when compared with national assessments. They are less likely to be attuned to local issues and concerns, and so to meet the needs of policy-makers, since compromise has to be reached between participating countries on the purposes, scope, and timing of an assessment. In particular, since compromise will be required in their construction, it is unlikely that

assessment instruments will reflect completely the curricular efforts of any individual country.

Not a great deal of information is available on the use that has been made of the findings of international studies. Most use seems to have involved the modification of intended curricula, though some impact on curricula as implemented in schools may be inferred if, as is claimed, changes occurred in the levels of achievement of Japanese students following changes in the intended curriculum. Some findings (relating, for example, to the role of the home in determining children's scholastic progress) point to the fact that some factors that affect students' learning may be ones that are not readily altered by the actions of policy-makers.

A number of global organizations (IEA, OECD) and a number of local consortia organize international assessments of achievement. A country considering participation in an assessment would need to establish contact with such a body. Participation in a regional assessment might be more attractive than participation in a worldwide assessment for a developing country, if for no other reason than that worldwide assessments have tended to be dominated by industrialized and economically developed countries. When the socio-economic and education conditions in countries participating in an assessment are broadly similar, the comparative data generated should be more meaningful and relevant to each individual country.

VII. Using student achievement data to evaluate schools

In this chapter, the use of information on student achievements derived from national and state assessments in which all schools participate to make judgements about the quality of individual schools (or teachers) is considered. Some consideration is also given to data obtained in public examinations since, in a number of countries (e.g. France, United Kingdom), these, as well as data derived from national assessments are used to construct 'league' tables of schools' performance. The consequences of making assessment information available in the form of league tables are described, as well as efforts to address the criticism that ranking schools, whether on the basis of state/national assessment or public examination performance, is unfair because it does not take account of how schools differ in the characteristics of their students and in the circumstances in which they operate.

Evaluating the quality of schools

There can be little doubt that most parents make judgements from time to time about the quality of individual schools. Judgements are likely to be based on their observations of a school's facilities and resources, their perceptions of the quality of teaching, perhaps the availability of extra-curricular activities, and, when external examinations are a feature of a system, parents are likely to have a good idea about how students perform. Such considerations may influence parents' choice of the school to which they will send their children, where choice exists. Parents have also been known to choose their place of residence for its proximity to a particular school.

In many countries, a school inspectorate is employed to assess the quality and effectiveness of individual schools and of teachers. To do this, inspectors normally observe teachers' classroom performance and assess student knowledge and skills in curriculum areas. Reports might be prepared on individual schools and teachers for ministry of

education records. Such systems of inspection are not a feature of all education systems and, even where they exist, their role has been diminishing as the power of teacher unions and recognition of the professional autonomy of teachers increase. In some countries, lack of resources renders inspection ineffective.

More formal procedures are now being used in a number of countries to arrive at judgements about the adequacy of the performance of an individual school, and even of individual teachers, based on the achievements of students. This may be done in the context of a state-wide assessment (e.g. in Brazil, the USA), of a national assessment (e.g. England, France, Nepal, Mauritius), or of an analysis of external (public) examination results (e.g. England, France, Kenya).

Action following the evaluation of schools

The kind of action taken at the individual school level on the basis of assessment information varies widely. Strengths and weaknesses of individual schools, based on a comparison of a school's own results with data for other schools, may be identified, as happens, for example, in national assessments in Chile, France, Thailand, and Uruguay. Suggestions may be made for addressing learning difficulties identified in the results, and schools may be encouraged to use the information for self-evaluation. In the Mendoza province in Argentina, the assessment is designed to influence teachers in their choice of the content and skills that are prioritized in class, and copies of assessment instruments are distributed so that teachers can model their classroom activities on the assessment (see *Box 7.1*). Schools that are performing poorly may be given additional resources, and may be assisted in identifying problems, setting targets, raising expectations, and developing appropriate strategies.

Box 7.1 Uses of results in Mendoza Province, Argentina

Following an assessment, each school received an individualized report at the beginning of the school year, which contained results for the school as well as provincial, departmental, and sectoral (urban, urban marginal, and rural) average scores. Test scores for the school were disaggregated by classroom and by student, as well as by skill area (geometry, probability, etc.) and test item. Schools were assigned an overall ranking relative to all other schools within their sector. A series of analytical documents, videos, and workshops about specific difficulties encountered by students proposing pedagogical changes to address perceived shortcomings was provided. Patterns of student achievement in terms of gender, geographical location, and public/private management variables were described.

Source: Benveniste, 2000b.

Some national and state assessments attempt to influence teacher behaviour by publishing the results of schools or by providing inducements for improved performance. Assessments that fall in the latter category are ones in which teachers receive additional remuneration if students achieve a target (if, for example, 85 per cent reach a satisfactory level of proficiency). Assessments that provide incentives (including money to teachers and schools) have been a feature of systems in Australia, Chile, and the USA (see *Box 7.2*).

In England, local authorities, on the basis of test results, provide support to schools in implementing an 'action plan' devised by the school to address problems. If results are particularly bad, the schools are given a deadline to improve and, if they fail, they can be closed or taken over.

Box 7.2 Chile: Index for merit awards for schools

Chile operates a cash reward system based on school-level performance. Schools compete with other schools in the region with similar geographic and socio-economic characteristics. The best-performing schools receive a monetary award equivalent almost to a monthly salary per teacher. Awards are based on scores on an index summarized below:

Variables	Indicators	Weight 1998-1999 %
Effectiveness	**SIMCE scores in mathematics and Spanish**	**37**
Value added	**Average SIMCE gain score**	**28**
Capacity for initiative	• Creation of teacher councils • Participation in monthly teacher-centre meetings • Pedagogical activities • Student council • School development plan • Teacher workshops	6
Improvements in working conditions	• Full staffing • Replacement for absent teachers	2
Equality of opportunities	• Student retention rate • Student graduation rate • Differential groupings • Integration projects • Absence of discriminatory practices	22
Integration of teachers and parents	Acceptance of educational work by parents, guardians and students, and creation of parent centres	5

While the overall index of school effectiveness encompasses a broad range of variables, educational achievement accounts for almost two-thirds of the overall index mark.

Source: Delannoy, 2000.

In Chile, 900 schools (about 10 per cent of schools in the country) are identified as being in need of assistance on the basis of their performance in language and mathematics assessments. Efforts are then made to improve performance in a variety of ways: by providing textbooks, classroom libraries, and pedagogical materials, and by arranging for teacher professional development through school-based workshops. Provision is made for after-school activities for students. There is evidence that the achievement gap between the schools and other schools diminished over time (Benveniste, 2000b).

League tables

In some countries, the results of students' performance on an assessment are published in a league table, in which schools are ranked in terms of the proportion of students who reached target levels. The practice, which attracts much media and professional comment, seems to be perceived as a relatively inexpensive way of addressing problems of quality. In some cases, it seems to be expected that it will result in higher standards simply by engendering competition between schools, teachers, and students (see Guskey and Kifer, 1990). It might, for example, serve to create a 'market' situation in which parents 'shopped around' for schools for their children, armed with 'evidence' about their performance.

There can be little doubt that when assessment data are published that show how well or how poorly schools are doing, this can impact on the behaviour of the schools. However, several factors indicate that the use of league tables is more complicated than might appear at first sight, is far from being free of problems, and may even be misleading. First, the performance of schools (and thus their position relative to the performance of other schools) may vary depending on the school outcome that is used. Thus, school rankings will differ depending on whether they are based on public examination results, on performance in basic curriculum areas, on an assessment by school inspectors, or on an overall measure of school effectiveness. Second, even rankings based on the *same* assessment can vary depending on the criterion of 'success' that is used. Thus, a ranking based on the proportion of students who obtain high grades (e.g. who on the basis

of their performance are classified as 'advanced') may differ from a ranking based on the proportion of students who achieve a more moderate but acceptable level of performance (e.g. are classified as 'proficient'). Third, the lack of precision in even well-constructed assessment instruments means that small differences between schools in scores on outcomes will be due to chance. A difference of a few points on an outcome score could result in a very large difference in a school ranking in a large school system. Fourth, whatever outcome measure is used, results can vary from year to year. This might be due to differences in cohorts of students, as well as to differences between teachers.

If one should be held accountable only for matters over which one has control (see *Chapter II*), then a further issue arises from the fact that student achievements are dependent on a variety of influences other than those brought to bear by the school (see *Box 7.3*). If the intention is to make a judgement about a school's contribution to student achievements, then the problem of how to make adequate allowance for factors outside the control of the school has to be addressed. Thus, it is reasonable to ask if students performed poorly because they were not motivated or had little home support, or because the school was inadequately resourced, or because teachers taught badly. Furthermore, if students differ from school to school in their level of achievements when entering a school, then a measure of achievement at a later date that does not take this into account may not adequately reflect a school's success in moving them from their initial entry level to their present level of achievement (see Goldstein and Thomas, 1996; Torrance, 1997).

Box 7.3 Factors that affect student achievement

- *Characteristics of students* when they enter the school in which they are assessed. These include genetic endowment, scholastic achievements, health and physical condition, aspirations, and self-image.
- *Conditions in which students live*. These include the physical conditions and resources of their homes and communities, as well as the support, guidance, and role models that are available to the student at home and in the community, some of which will depend on family income, parents' educational level, attitudes, expectations, and aspirations.
- *School conditions*. These are threefold.
- Some conditions are under the control of the school: what teachers and other adults in the school do to help students learn in formal lessons, and what is provided by way of sport, consultation with parents, and pastoral care.
- Other school conditions are not normally under the control of the school. These include the space and other resources that are available.
- Some school conditions may or may not be under the control of the school. For example, characteristics of the study body (which can impact on individual students' achievements and aspirations) may be manipulated by the school through selection procedures. In some education systems, this option is not available.

Measuring the school's contribution to students' achievement

To address the criticism that comparisons may be unfair because schools differ in the characteristics of their students and the circumstances in which they operate, one of a number of approaches may be adopted. In the first, comparisons are confined to schools that are similar in the socio-economic characteristics of their students. For example, in England, schools are placed in one of five categories on the basis of the percentage of their students who are eligible for free meals. They are then encouraged to compare the performance of their students on an assessment with the performance of students in schools that have been assigned to the category in which they have been placed. In this approach, eligibility for free meals is used as a proxy for the second factor (conditions in which students live) in *Box 7.3*.

In France, the reporting of the results of the school-leaving examination (*Baccalauréat*) goes a step further in an attempt to contextualize students' performance and to take into account factors other than teaching in the school that may have affected it. The step involves providing what is called 'value added' information. This is done by calculating the expected success rate on the examination when the socio-economic background of students (categorized as 'very advantaged', 'advantaged', 'medium', and 'disadvantaged') has been taken into account. Schools are provided with two sets of data on the examination: the actual success rate of students in the school, and the rate that would be expected for a school with its particular socio-economic characteristics. A school in which the actual success rate exceeds the expected rate is considered effective in adding 'value' to students' achievements, while one in which the expected rate exceeds the actual rate is considered to be doing poorly.

In this case, there is in fact no real measure of the value that has been added since no data are provided on the base that was there to begin with. To address this issue, some systems (e.g. in some American states and in some local education authorities in the United Kingdom) obtain data on students' achievements at an earlier point in time to obtain an estimate of students' 'progress' from that base. This may be combined with other information (e.g. students' socio-economic background and ethnicity) to calculate students' expected achievements. Again, the difference between students' actual and expected achievements is used to estimate the 'value' that schools have added.

Different rankings emerge when school performances are compared using unadjusted performance scores, scores adjusted for contextual factors (e.g. the percentage of students entitled to free school meals), and scores adjusted for earlier achievement. Thus, a school that merits a very high rank on the basis of its raw scores may turn out to be 'average' when account is taken of the prior achievements of its students. On the other hand, a school that merits a low rank on the basis of its raw scores may be 'average' when account is taken of the proportion of its students that quality for free meals or of the students' achievements at an earlier point in time (Goldstein, 2001). Thus the importance of taking into account the

conditions in which schools operate in attempting to estimate their effectiveness is underlined.

Adjusted estimates of student achievement may have value in identifying very discrepant schools in which further investigation with a view to providing some assistance to them would seem appropriate. However, the limitations of even such estimates in judging the effectiveness of schools should not be overlooked. These include the fact that the range of achievements assessed is usually limited; that data are often incomplete because of student absenteeism or student turnover rates; that the many personal, school, and background factors that can affect achievement are inadequately specified (not all of the factors listed in *Box 7.3* are taken into account); that schools may achieve quite different results for initially low and initially high achieving students; and that outcome measures will be unreliable when the number of students in a school is small.

The impact of national assessment data on schools

A variety of effects of national assessments on schools have been identified. These are most pronounced when results are placed in a high-stakes context, that is, when information on the performance of schools is published, or when sanctions (for schools, teachers, or students) are attached to performance.

Some impact may be achieved simply through participation in an assessment. For example, the British assessment of the national curriculum set out to affect how teachers taught and assessed their students, and, it would seem, achieved some success in prompting them to reflect more on the curriculum content to be taught and on the educational process. In Thailand, measures of affective outcomes, practical skills, and social perception were included in a national assessment in an effort to dislodge teachers' preoccupation with cognitive development in specific content areas (a preoccupation that was reinforced by end-of-school examinations for university entrance). This led teachers to place greater emphasis on these outcomes in their teaching and assessments (Pravalpruk, 1996).

A number of positive effects have been reported for state-wide assessments in which results are published in the USA. For example, states in which dissemination of test scores was required registered a slight increase in the number of students in academic tracks over the years compared to ones which did not have this requirement. There is also some evidence that holding schools accountable helped reduce teacher bias, raise their expectations for student performance, and increase high-school graduation rates (Scheurich, Skrla and Johnson, 2000).

Negative consequences have also been identified when high stakes (in the form of some kind of sanction) are attached to a school's performance in a national or state assessment. As has been documented in the case of external (public) examinations, students will be coached on the assessment instruments, and teachers will focus their efforts on what is measured, leading to a narrowing of the curriculum and concentration on achieving high levels of test performance at the expense of general cognitive development and deeper forms of learning (Kellaghan and Greaney, 1992; Kellaghan, Madaus and Raczek, 1996; Madaus, 1988). Adequate attention may not be given to other aspects of student achievement, for example, growth in social awareness and emotional development. There is also evidence that national and state assessments, even when stakes are relatively low, are associated with cheating behaviour by students, teachers, and school administrators.

A further consequence of attaching high stakes to performance is that, in an effort to improve the mean level of performance of a school, teaching resources and strategies may be focused on 'borderline' students (to increase the number classified as proficient), while lower and higher achieving students may be neglected. Schools may also adopt procedures to reduce the number of students of low achievement who sit the assessment tasks. This may involve a number of strategies: being more selective in the students that the school enrols, retaining students in grades in which students are not assessed, or increasing the number of students classified as having a disability if regulations permit their exclusion from the assessment. Each of these strategies will reduce the number of students with low levels of achievement taking an assessment, and this, of course, will impact on the school's average performance.

Examples of the use of such strategies are not difficult to find. The introduction of the Grade 10 assessment in Texas in the USA has been associated with an increase in the Grade 9 retention rate, in the school drop-out rate, and in the time devoted to subjects tested at the expense of other subjects (Haney, 2000). Furthermore, assessment was associated with an increase in the number of students described as in need of special educational provision, who were exempted from the assessments. A rather different approach to the manipulation of an assessment is to be found in Chile's special programme for low-performing schools. In this case, teachers underrated the socio-economic status of their students, realizing that this would affect the interpretation of results (Himmel, 1996).

Parents may also react to high-stakes assessment results, choosing schools that they believe will provide the highest level of achievement for their children, thus compounding the difficulties of some schools and their students. Finally, high-stakes testing has been associated with problems in the recruitment and retention of teachers. Teachers leave schools in which student performance is poor, and are attracted to schools in which they will receive a bonus for good performance.

Conclusion

The primary purpose of a national assessment is to obtain information that will be of value to policy-makers and managers in making decisions about the allocation of resources in the education system. The information should indicate the level of achievement of students in the system at particular points of interest and, depending on its design, the relative performance of students in varying components of achievement as well as the achievements of sub-groups in the population. The results of some state and national assessments, however, are also used to evaluate individual schools and to have an immediate impact on teachers' behaviour in the classroom.

If the decision is to use an assessment for the evaluation of schools, it also has to be decided whether or not sanctions will be attached to performance for individual schools, teachers, or even students. If the decision is not to attach sanctions, it may be assumed that the education authority accepts major responsibility for addressing any deficiencies

BEST COPY AVAILABLE

that may be identified. If, on the other hand, the attachment of sanctions to performance is proposed, a number of considerations merit attention.

First, it should be recognized that school improvement is a dynamic complex process which requires a broad range of information and of action. The information derived from a single assessment will not on its own meet this need. It does not tell us how and why a school was 'ineffective', or how problems that may exist should be addressed. All it does is provide a starting point for further investigation which might take the form of school visits, interviews, and observation to analyze the situation and to devise strategies to address deficiencies. Second, the accountability of schools should be considered in the context of other institutions, agencies, and individuals that also are accountable in some way for the operation of schools and the resources that are allocated to them. Third, in assessing the performance of schools, it is necessary to balance the need for political accountability with a recognition of the professional autonomy of teachers. A heavy emphasis on the former can easily result in a standardized and mechanical approach to teaching.

Fourth, school performance should be assessed in the context of the total operation of the school: its mission, organization, activities and all its curricular aspirations, not just student achievement. Fifth, the use of assessment data to evaluate schools may give rise to serious equity issues if due allowance is not made for the fact that schools differ greatly in the characteristics of the students that attend them. It also may be a disservice to schools that work in difficult environments and may be successful in areas of achievement other than those assessed. In making comparisons between schools, efforts should be made to take into account the circumstances in which they operate, while at the same time recognizing that it is impossible to specify and make allowance for all such circumstances. Sixth, schools that score within a narrow range on an assessment task should not be regarded as differing 'significantly' from each other on the outcome that has been assessed. Small differences will inevitably arise from errors in measurement. Finally, it should be recognized that attaching high stakes to performance in the form of sanctions for schools, teachers, or students may result in unsound pedagogical practices, various forms of malpractice, and a 'teaching-to-the-test' culture.

VIII. Issues in the design of an assessment

In *Chapter I* it was stated, perhaps in over-simplistic terms, that education policy-makers, planners, and managers would need to answer two basic questions in considering whether or not to invest in the use of assessment to improve the quality of education. Should resources be invested in assessment activities? And if the answer is yes, what kind of assessment and what uses of assessment information are likely to impact on the quality of education in schools?

In answer to the first question, there can be little doubt that there is a need to improve assessment practices in external (public) examinations and in everyday teaching in classrooms. Deficiencies in both have been identified, and it may reasonably be expected that efforts to address them would have positive effects on the quality of students' learning. Addressing deficiencies in external (public) examinations, where such examinations exist, would require only a marginal increase in expenditure, since a considerable amount of money is already spent on them. If this is accepted, then the question to be addressed by decision-makers can be narrowed to one in which the choice is between increasing investment in teachers' classroom assessment skills or investing in new forms of assessment represented in national and international assessments.

If the only aim is to improve the quality of students' learning, and if cost is not a consideration, then the decision would have to come down in favour of improving teachers' classroom assessment skills. Indeed, this is the only category of assessment that has as its primary and immediate objective the facilitation of student learning. While other categories of assessment may aspire to affect learning, they can hope to do so only indirectly. Unfortunately, improving the assessment skills of teachers is also likely to be the most expensive option.

However, the improvement of student learning is not the only purpose of assessment (see *Chapter II*). If education policy-makers, planners, and managers need information about the adequacy of the

performance of an education system, or of its components, that need can only be met by a national assessment. While some information on student achievements in education systems can, and often is, obtained from other sources, it cannot be regarded as equivalent to that obtained in a national assessment. Thus, a system of supervision or inspection can provide important information on student achievements, but it will not be systematic or comparable across schools. While an analysis of public examination results can also provide information on student achievements, such results lack the basis for comparability that is required for monitoring over time since, on the one hand, examination populations change in an unknown way and, on the other hand, standards cannot be regarded as consistent across years (Kellaghan, 1996a).

The rapid spread of national assessment activity throughout the world, particularly in the 1990s, often in the context of a societal emphasis on indicators, accountability, and transparency, would seem to underline the need for objective and systematic information about the quality of student learning at several planning and decision-making levels in the education system that is not otherwise available. To what extent the information might also be used to improve the quality of students' learning is of course another matter. Whether or not it will prove useful for this purpose would seem to depend on a number of factors: the quality of the assessment; the action that is taken to communicate results to those who can act on them (national, regional, and local administrators; curriculum developers; teachers; and teacher educators); and the extent to which the assessment is aligned with other instructional guidance mechanisms in the education system.

If a decision is made to undertake a national assessment, then the need to make a series of further decisions arises. These decisions are outlined in the next section. The decisions to be made in an international assessment are similar but will, of course, require the agreement of participating countries. A brief description of the advantages and disadvantages of participating in an international assessment is followed by a consideration of the question of cost. Finally, issues that arise in the use of information derived from an assessment are considered.

Considerations in designing an assessment

Following a decision to undertake a national assessment, a decision has to be made about the curriculum area to be assessed. Going on general practice in national/international assessments, information on achievement in core curriculum areas would seem to be the primary need. While basic literacy and numeracy skills are, of course, of crucial importance, they are not the only aspects of student development that schools attempt to influence. Other cognitive and non-cognitive skills are also important and, when included in an assessment, may induce teachers to pay more attention to them in their teaching. Information on students' attitudes, activities, interests, and values, although difficult to assess reliably, may also be useful in interpreting findings on achievement.

A decision will also be required about the form of assessment that will be employed. The choice is between a complex procedure extending over several days involving observation, tests, and checklists carried out by students' own teachers (in the United Kingdom) or by teachers from another school (in New Zealand), or a test administered under standardized conditions, which is the procedure in most countries. The former is proposed as providing a more 'authentic' assessment of students' performance in a variety of situations, and is designed to impact directly on teachers' assessment practices and to influence their choice of curriculum topics; the latter is regarded as more 'objective' and as providing a better basis for comparisons between schools, or over time.

The quality of the instruments used to measure achievement merits more attention than it is given. While the instruments used in international studies tend to be of a high technical quality, the same cannot be said of the instruments used in all national assessments. A need for improvement is indicated even in the case of instruments designed to measure basic scholastic competences. There is also a need in the case of instruments used in the assessment of older students to ascertain if they assess the high-level skills and knowledge that it is claimed students require, or if the skills that are assessed have any relevance for life in a global competitive economy.

A further decision relates to the 'population' that will be targeted in an assessment. Unless it is decided to include all students in all grades in a school, it has to be decided whether the basis of inclusion will be students' age or the grade level at which they are enrolled. The selection of grade rather than age has the advantage that it is easier to administer an assessment to students who are in the same grade than to students spread across a number of grades, and in education systems in which most students of a particular age are enrolled in the same grade, there will not be much of a problem. However, in systems in which children vary in the age at which they enrol in school, and/or where retention in grades is common, there may be an interest in obtaining achievement data relating to age rather than, or in addition to, data relating to grade level.

If practice in most countries is followed, then the particular age(s) or grade levels(s) selected for a national assessment will fall within the period of compulsory education (most likely at primary-school level); furthermore, students will have been in school long enough for education to have had an impact. More than one grade level may be selected to provide some indication of change associated with years of schooling (e.g. from Grades 3 to 5), though if this is planned, some of the items in the assessment instruments should be common to both levels.

A further decision is whether the assessment exercise is to be carried out in all schools, or if inferences about the performance of the education system are to be based on an assessment carried out in a sample of schools. There is no point in adopting the former course, which will be very much more expensive, unless plans are in place to act on information obtained in the assessment at the individual school level.

A decision also has to be taken about the data that will be collected in addition to achievement data (e.g. in questionnaires completed by teachers, students, and perhaps parents) which may help in the interpretation of findings regarding students' achievements. It is usual, for example, to collect information about inputs, such as buildings, teacher qualifications, and the socio-economic conditions of students'

families and communities. Information is also sometimes obtained on schools' curricular and instructional activities. While it may be tempting to collect as much of this kind of information as possible, serious consideration should be given in advance to how data will be used and to the availability of resources to carry out analyses. It should also be appreciated that some issues of interest might be more precisely and less expensively addressed in other types of research.

It is important when planning an assessment to have in mind the information needs of a variety of publics, as these can have implications for the design of instruments and the size of samples. It may, for example, be proposed to analyze data by curriculum domain, providing evidence that is likely to be of value to education managers, curriculum developers, textbook writers, and teachers, on strengths and weaknesses of students (in, for example, their problem-solving ability in mathematics or their ability to integrate information or draw inferences in reading). If so, the design of an instrument should be such that it yields the type of information that is required. If information on sub-populations (e.g. on regional differences) is required – information that is likely to be of interest to politicians, planners, and those involved in resource allocation – then this will have implications for the design and size of the sample that will be required.

Before an assessment is carried out, serious consideration should also be given to the resources, political and technical, that will be required to support it and see it through. At the political level, the involvement of a wide range of stakeholders (for example, in a steering committee) can help ensure that the concerns of all involved in the education system are represented, as well as providing support for the exercise. At the technical level, competence in assessment instrument construction, sampling, analysis, and the management of a large-scale survey will be required.

Conditions in many developing countries may make it difficult to carry out an assessment. At the administrative level, there may be inadequate funding, a lack of technical capacity, incomplete data on schools, and logistical problems associated with a poor infrastructure in transport, telecommunications, and postal services. Linguistic

diversity may also need to be addressed. Nigeria's EFA assessment was confined to Grade 4, as any lower level would have required translating assessment tasks into 270 languages. The tradition of grade repetition, which is particularly prevalent in francophone African and Latin American countries, poses sampling and logistical difficulties. At the school level, problems may arise from overcrowded classrooms, inadequate seating and conditions for administering assessment tasks, and teacher and student absenteeism.

Participation in an international assessment

An international assessment is less likely than a national assessment to be focused on the needs of a particular education system. Furthermore, the inevitable demands made by an international study to meet deadlines may pose problems for some countries. However, these facts should be weighed against some of the advantages of participation. Development and analytic costs may be less because they are shared with other countries, while the capacity of a national education system may be enhanced through interaction with individuals who have skills in assessment in other education systems. An international assessment also provides the opportunity of reviewing curricula in light of curricula in other countries.

The political dimension to participation in an international assessment should not be overlooked. Some countries may not wish to participate, given the likelihood that their students may perform poorly, which in turn might cause political embarrassment. While there are advantages in participating in an assessment in which neighbouring countries are involved, a tradition of antagonism between them may lead to a decision not to, given the likely political fallout.

Cost

The question of cost is one that must inevitably be addressed in considering a national assessment or participation in an international assessment. It is not possible to estimate the precise cost benefits of various types of assessment in improving the quality of education that might guide planners and administrators in making decisions about

the allocation of resources. While costs could be calculated, associated benefits cannot be accurately identified. This does not mean that consideration of cost and of expected benefits should not figure in the decision-making process. At the very least, a judgement is required about the likely value of an assessment in aiding decision-making about the use of resources, for example, by identifying areas of weakness in a curriculum or inequities in the education system. Information that leads to improved curricula or textbooks could more than cover the cost of obtaining it.

The cost of an assessment may seem high when considered in its totality. However, it may seem much more attractive when the cost per student in the education system is calculated. Furthermore, a number of decisions can help to contain costs. Location of the assessment operation in a research organization, where one exists, for example, has the advantage that it makes use of existing relevant skills (e.g. for instrument design, sampling, analysis). Some countries which do not have such a facility have located their national assessment in an examinations authority. In this case, investment may be required to provide individuals with relevant skills (particularly in relation to sampling and analysis). Expertise may also be available in a census bureau (for assistance in sampling) and among teacher educators and the inspectorate (for curriculum analysis).

The use of information derived from an assessment

Given the very considerable resources invested in national and international assessments, some additional expenditure to ensure that the information they provide is fed back into the education system would seem to be a cost-effective procedure. While there is some evidence that data from assessments have been used for policy decisions (for example, to alter curricula or to address equity issues), it is likely that the potential of the data is not being fully exploited.

There is abundant evidence from many countries in which assessments have been administered that the information derived from them is not being effectively communicated, and so is not being integrated into decision-making structures in ministries of education

or at other levels of the education system. In this situation, it would be unrealistic to expect assessments to have much impact on the quality of students' learning. The communication of information may require a number of strategies. While it may be possible to provide all the information derived from an assessment in a single report, it may be more effective and user-friendly to prepare a number of reports. Summary data may be adequate in a publication prepared for the public. Policy-makers will require more information (e.g. details of the performance of sub-groups in the population), while curriculum analysts, textbook writers, and teachers will need to know how students performed in particular areas of the curriculum. A description of technical aspects of an assessment (e.g. the assessment instrument, method of sampling, rules of exclusion) will be particularly important if the assessment is to be repeated at a later date.

Getting results in a meaningful form to teachers poses particular challenges. There is not a great deal of experience to go on in doing this, of the problems that it may give rise to, or of the impact it might have. Printed summaries, charts that display differences by curriculum domain, and workshops have been used in some countries. Continuing support following the initial communication has also been necessary. In communicating information, account should be taken of conditions in schools, such as inadequate curricula and textbooks, the influence of external examinations, large class sizes, and teachers' competence, any of which may inhibit changes in teaching style.

A particular problem in some national and state (though not international) assessments arises when results are used to perform functions that previously were performed by other methods of assessment. For example, results may be used to evaluate schools (a traditional function of a school inspectorate), or to make decisions about student promotion or graduation (which, in the past, was done on the basis of an external or public examination in some countries, and by teachers in other countries). When national assessment results are used for these purposes, issues relating to accountability and unintended consequences arise. Holding schools primarily accountable for students' achievements is problematic for at least three reasons. First, it does not recognize that teachers are not the only individuals

BEST COPY AVAILABLE

that are accountable for student performance. Second, it may not make adequate allowance for the circumstances in which schools operate or for the characteristics of their students. Attempts to do this have not been entirely satisfactory. Third, rather than improving the quality of students' learning, the use of assessment results to hold schools accountable may lead to the adoption of a variety of strategies and pedagogical practices (such as 'teaching to the test') to improve the mean level of a school's performance on the assessment instruments, an improvement which, however, may not be matched in the broader achievement constructs that a reform was intended to promote.

Conclusion

The considerable efforts in the 1980s and 1990s throughout the world to develop procedures to describe the achievements of education systems, and to develop capacity to implement the procedures, may be taken as evidence of a perceived need to obtain objective and systematic information on students' learning. During this time, much progress was made in development of the techniques required for an assessment, particularly ones relating to the construction of assessment instruments and sampling, and in the transfer of the knowledge required to carry out an assessment. A reasonably well-developed technology exists to describe how well students are learning in the education system, to identify sub-groups that may be performing poorly, and perhaps to a lesser extent to determine if student achievements change over time. A range of options is now available to education policy-makers, planners, and managers who wish to add information on student learning to the information they already have on the resources of education systems.

That is not to say that techniques for carrying out assessments are perfect, or that the quality of all the national assessments that have been carried out can be regarded as satisfactory. Considerable room for improvement remains in the design of assessment instruments, and problems have arisen in how results are presented, communicated, and used. Those responsible for evaluation policy in ministries of education should be aware of the limitations that this situation gives

rise to. They should also bear in mind that assessment takes many forms and serves several purposes; thus, commitment to national assessment should not lead to a neglect of the need to improve other forms of assessment, including the system of supervision and school support, teachers' practices in their own classrooms, and external (public) examinations. Policy-makers should also strive to ensure that national assessments are aligned with other major instructional guidance mechanisms in the education system, not only with other assessment systems, but with curricula, teacher education, school capacity building, and measures to address inequities. In the absence of such alignment, a national assessment is unlikely to be an effective instrument in improving the quality of education.

References

AFT (American Federation of Teachers); NCME (National Council on Measurement in Education); NEA (National Education Association). 1990. "Standards for teacher competence in educational assessment of students." In: *Educational Measurement: Issues and Practices, 9*(4), pp.30-32.

Baker, E.L.; Linn, R.L. 1995. "United States." In: *Performance standards in education. In search of quality*. Paris: OECD.

Beaton, A.E.; Postlethwaite, T.N.; Ross, K.N.; Spearritt, D.; Wolf, R.M. 1999. *The benefits and limitations of international educational achievement studies*. Paris: UNESCO/International Institute for Educational Planning; International Academy of Education.

Benveniste, L. 2000*a*. "Student assessment as a political construction: The case of Uruguay." In: *Education Policy Analysis Archives, 8*(32).

Benveniste, L. 2000*b*. *The politics of student testing: a comparative analysis of national assessment systems in Southern Cone countries*. Doctoral dissertation, Stanford University.

Burnstein, L.; Oakes, J.; Guiton, G. 1992. "Education indicators." In: Alkin, M.C. (Ed.), *Encyclopedia of educational research* (6th ed.). New York: Macmillan.

Carroll, D. 1996. "The grade 3 and 5 assessment in Egypt." In: Murphy, P. et al. (Eds.), *National assessments: testing the system*. Washington DC: World Bank.

Chinapah, V. 1997. *Handbook on monitoring learning achievement. Towards capacity building*. Paris: UNESCO.

Cosgrove, J.; Kellaghan, T.; Forde, P.; Morgan, M. 2000. *The 1998 National Assessment of English Reading*. Dublin: Educational Research Centre.

Crespo, M.; Soares, J.F.; de Mello e Souza, A. 2000. "The Brazilian national evaluation system of basic education: context, process, and impact". In: *Studies in Educational Evaluation, 26*, pp.105-125.

Daugherty, R. 1997. "National curriculum assessment: the experience of England and Wales". In: *Educational Administration Quarterly, 33*, pp.198-218.

Delannoy, F. 2000. *Education reforms in Chile 1980-98: a lesson in pragmatism*. Washington DC: World Bank.

Flockton, L.; Crooks, T. 1998. *Social studies: assessment results, 1997. National Education Monitoring Project Report 8*. Dunedin, New Zealand: Educational Assessment Research Unit. University of Otago; Ministry of Education.

Goldstein, H. 1996. "International comparisons of student achievement." In: Little, A. and Wolf, A. (Eds.), *Assessment in transition. Learning, monitoring and selection in international perspective*. Oxford: Pergamon.

Goldstein, H. 2001. "Using pupil performance data for judging schools and teachers: scope and limitations." In: *British Educational Research Journal, 27*, pp.433-442.

Goldstein, H.; Thomas, S. 1996. "Using examination results as indicators of school and college performance." In: *Journal of the Royal Statistical Society. 159*, pt.1, pp.149-163.

Greaney, V.; Kellaghan, T. 1996. *Monitoring the learning outcomes of education systems*. Washington DC: World Bank.

Greaney, V.; Khandker, S.R.; Alam, M. 1998. *Bangladesh. Assessing basic learning skills*. Dhaka: University Press; World Bank.

Guskey, T.R.; Kifer, E.W. 1990. "Ranking school districts on the basis of statewide test results: is it meaningful or misleading?" In: *Educational Measurement. Issues and Practice, 9*(1), pp.11-16.

Haney, W. 2000. "The myth of the Texas miracle in education." In: *Education Policy Analysis Archives, 8*(41).

Haney, W.; Raczek, A. 1994. *Surmounting outcomes accountability in education*. Paper prepared for US Congress Office of Technology Assessment. Chestnut Hill MA: Center for the Study of Testing, Evaluation and Educational Policy, Boston College.

Himmel, E. 1996. "National assessment in Chile." In: Murphy, P. et al. (Eds.), *National assessments. Testing the system*. Washington DC: World Bank.

Howie, S.J. 2001. *Mathematics and science performance in grade 8 in South Africa 1998/1999*. Pretoria: Human Sciences Research Council.

Husén, T.; Postlethwaite, T.N. 1996. "A brief history of the International Association for the Evaluation of Educational Achievement (IEA)." In: *Assessment in Education, 3*, pp.129-141.

Johnson, E.G. 1992. "The design of the National Assessment of Educational Progress." In: *Journal of Educational Measurement, 29*, pp.95-110.

Kane, M.T. 2001. "So much remains the same: conception and status of validation in setting standards." In: Cizek, G.J., *Setting performance standards. Concepts, methods, and perspectives*. Mahwah NJ: Lawrence Erlbaum.

Kellaghan, T. 1996*a*. "Can public examinations be used to provide information for national assessment?" In: Murphy, P. et al. (Eds.), *National assessments. Testing the system*. Washington DC: World Bank.

Kellaghan, T. 1996*b*. "IEA studies and educational policy." In: *Assessment in Education, 3*, pp.143-160.

Kellaghan, T. 1997. "Seguimiento de los resultados educativos nacionales." In: Álvarez H. B. and Ruiz-Casares, M. (Eds.), *Evaluación y reforma educativa. Opciones de política*. Washington DC: US Agency for International Development.

Kellaghan, T.; Greaney, V. 1992. *Using examinations to improve education. A study in fourteen African countries*. Washington DC: World Bank.

Kellaghan, T.; Greaney, V. 2001. "The globalisation of assessment in the 20th century." In: *Assessment in Education, 8*, pp.87-102.

Kellaghan, T.; Grisay, A. 1995. "International comparisons of student achievement: problems and prospects." In: *Measuring what students learn*. Paris: OECD.

Kellaghan, T.; Madaus, G.F.; Raczek, A. 1996. *The use of external examinations to improve student motivation*. Washington DC: American Educational Research Association.

Kelly, M.J.; Kanyika, J. 2000. *Learning achievement at the middle basic level. Summary report on Zambia's national assessment project 1999*. Lusaka: Examinations Council of Zambia.

Kirst, M.W. 2000. "Accountability: implications for state and local policymakers." In: Stufflebeam, D.L.; Madaus, G.F.; Kellaghan, T. (Eds.), *Evaluation models: viewpoints on educational and human services evaluation* (2nd ed.). Boston: Kluwer Academic.

Linn, R.L. 1983. "Testing and instruction: Links and distinctions." In: *Journal of Educational Measurement, 20*, pp.179-189.

Linn, R.L. 2000. "Assessments and accountability." In: *Educational Researcher, 29*(2), pp. 4-16.

Machingaidze, T.; Pfukani, P.; Shumba, S. 1998. *The quality of education: some policy suggestions based on a survey of schools in Zimbabwe*. Paris: UNESCO/International Institute for Educational Planning.

Madaus, G.F. 1988. "The influence of testing on the curriculum." In: Tanner, L.N. (Ed.), *Critical issues in curriculum*. Eighty-seventh Yearbook of the National Society for the Study of Education, Part I. Chicago: University of Chicago Press.

Madaus, G.F.; Kellaghan, T. 1992. "Curriculum evaluation and assessment." In: Jackson, P. (Ed.), *Handbook of research on curriculum*. New York: Macmillan.

Michel, A. 1995. "France." In: *Performance standards in education. In search of quality*. Paris: OECD.

Murphy, P.; Greaney, V.; Lockheed, M.E.; Rojas, C. (Eds.) 1996. *National assessments. Testing the system*. Washington DC: World Bank.

OECD (Organisation for Economic Co-operation and Development). 2001. *Knowledge and skills for life. First results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD.

Palestine Ministry of Education. 1998. *Level of achievement in Arabic for sixth-grade students at the end of the lower basic stage in Palestine*. Ramallah: Palestine Ministry of Education.

Popham, W.J. 1987. "The merits of measurement-driven instruction." In: *Phi Delta Kappan, 68*, pp.679-682.

Pravalpruk, K. 1996. "National assessment in Thailand." In: Murphy, P. et al. (Eds.), *National assessments. Testing the system*. Washington DC: World Bank.

Robitaille, D.F.; Beaton, A.E.; Plomp, T. (Eds.). 2000. *The impact of TIMSS on the teaching and learning of mathematics and science*. Vancouver: Pacific Educational Press.

Rojas, C.C.; Esquivel, J.M. 1998. *Los sistemas de medición del logro académico en Latinoamérica*. Washington DC: World Bank.

Ross, K.N. (Ed.) et al. 2000. "Translating educational assessment findings into educational policy and reform measures: lessons from the SACMEQ initiative in Africa." Paper presented at World Education Forum, Dakar, Senegal, 26-28 April.

Ross, K.N.; Mählck, L. (Eds.). 1990. *Planning the quality of education. The collection and use of data for informed decision-making*. Paris: UNESCO/International Institute for Educational Planning.

Scheurich, J.J.; Skrla, L.; Johnson, J.E. 2000. "Thinking carefully about equity and accountability." In: *Phi Delta Kappan, 82*, pp. 293-299.

Shore, A.; Pedulla, J.; Clarke, M. 2001. *The building blocks of state testing programs*. Chestnut Hill MA: National Board on Educational Testing and Public Policy, Boston College.

Torrance, H. 1997. "Assessment, accountability, and standards: using assessment to control the reform of schooling." In: Halsey, A.H. et al. (Eds.), *Education, culture, economy, and society*. Oxford: Oxford University Press.

UNESCO. 2000. *The Dakar Framework for Action. Education for all: meeting our collective commitments*. Text adopted by the World Education Forum, Dakar, Senegal, 26-28 April. Paris: UNESCO.

World Declaration on Education for All. 1990. Adopted by the World Conference on Education for All, Meeting Basic Learning Needs. New York: UNDP; UNESCO; UNICEF; World Bank.

IIEP publications and documents

More than 1,200 titles on all aspects of educational planning have been published by the International Institute for Educational Planning. A comprehensive catalogue is available in the following subject categories:

Educational planning and global issues
General studies - global/developmental issues

Administration and management of education
Decentralization - participation - distance education - school mapping - teachers

Economics of education
Costs and financing - employment - international co-operation

Quality of education
Evaluation - innovation - supervision

Different levels of formal education
Primary to higher education

Alternative strategies for education
Lifelong education - non-formal education - disadvantaged groups - gender education

Copies of the Catalogue may be obtained on request from:
IIEP, Dissemination of Publications
information@iiep.unesco.org
Titles of new publications and abstracts may be consulted at the following Web site: *http://www.unesco.org/iiep*

Imprimerie Alençonnaise
2, rue Édouard-Belin, 61002 Alençon
Dépôt légal : 2e trimestre 2002 - N° d'ordre : 50289

The booklet

Who would not wish to know if students, as a result of their exposure to schooling, are acquiring appropriate knowledge, skills, behaviour, and attitudes? And who would not wish to have this information if it could play an important role in improving the quality of education? National and international assessments, by specifying procedures to use data on individual students to describe the achievements of a whole education system (or a clearly defined part of it) hold out the promise of providing this information. It is not surprising, then, that the number of countries that have embarked on these assessments increased dramatically in the 1990s.

In this booklet, the nature, models, and purposes of national assessments, and how information derived from them has been used in countries across the world, are described. International assessments, which share many procedural features with national assessments, but are designed to allow comparisons across countries, are also described.

The booklet should be of interest to education policy-makers, planners, and managers who wish to add information on student learning to the information they already have on inputs to their education systems. For those who have already carried out an assessment, the booklet provides the opportunity to evaluate it in light of practice in other countries.

The author

Thomas Kellaghan is Director of the Educational Research Centre at St Patrick's College in Dublin, and a Fellow of the International Academy of Education. He served as president of the International Association for Educational Assessment from 1997 to 2001. Areas of research include assessment, programme evaluation, and factors affecting human development.

Vincent Greaney is Lead Education Specialist at the World Bank, Washington DC. He has worked in the USA, Ireland, Africa, and Asia. Areas of professional interest include public examinations, national assessment, reading, teacher evaluation, and development education.

UNESCO

iiep
international institute
for educational planning

ISBN 92-803-1215-4
9 789280 312157

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

X	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").